

WO2083953

Publication Title:

METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

Abstract:

The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

Data supplied from the esp@cenet database - <http://ep.espacenet.com>

This Patent PDF Generated by Patent Fetcher(TM), a service of Stroke of Color, Inc.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number
WO 02/083953 A1

(51) International Patent Classification⁷: C12Q 1/68, (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(21) International Application Number: PCT/US02/11757

(22) International Filing Date: 11 April 2002 (11.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/282,965 11 April 2001 (11.04.2001) US

(71) Applicant (*for all designated States except US*): PTC THERAPEUTICS, INC. [US/US]; 100 Corporate Court, Middlesex Business Center, South Plainfield, NJ 07080 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): RANDO, Robert [US/US]; 3 Brown Court, Annandale, NJ 08801 (US). WELCH, Ellen [US/US]; 33 Hollow Brook Road, Califon, NJ 07830 (US).

(74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report
— with amended claims

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/083953 A1

(54) Title: METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

(57) Abstract: The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

5 This application claims the benefit of U.S. Provisional Application No.
60/282,965, filed April 11, 2001, which is incorporated herein by reference in its entirety.

1. INTRODUCTION

10 The present invention relates to a method for screening and identifying test
compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-
competitive binding assays are advantageously used to screen libraries of compounds for
those that selectively bind to a preselected target RNA. Binding of target RNA molecules to
a particular test compound is detected using any physical method that measures the altered
physical property of the target RNA bound to a test compound. The methods of the present
invention provide a simple, sensitive assay for high-throughput screening of libraries of
15 compounds to identify pharmaceutical leads.

2. BACKGROUND OF THE INVENTION

20 Protein-nucleic acid interactions are involved in many cellular functions,
including transcription, RNA splicing, mRNA decay, and mRNA translation. Readily
accessible synthetic molecules that can bind with high affinity to specific sequences of
single- or double-stranded nucleic acids have the potential to interfere with these
interactions in a controllable way, making them attractive tools for molecular biology and
medicine. Successful approaches for blocking function of target nucleic acids include using
25 duplex-forming antisense oligonucleotides (Miller, 1996, *Progress in Nucl. Acid Res. &*
Mol. Biol. 52:261-291; Ojwang & Rando, 1999, *Achieving antisense inhibition by*
oligodeoxynucleotides containing N, modified 2'-deoxyguanosine using tumor necrosis
factor receptor type 1, METHODS: A Companion to Methods in Enzymology 18:244-251)
and peptide nucleic acids ("PNA") (Nielsen, 1999, *Current Opinion in Biotechnology*
30 10:71-75), which bind to nucleic acids via Watson-Crick base-pairing. Triplex-forming
anti-gene oligonucleotides can also be designed (Ping *et al.*, 1997, *RNA* 3:850-860;
Aggarwal *et al.*, 1996, *Cancer Res.* 56:5156-5164; U.S. Patent No. 5,650,316), as well as
35 pyrrole-imidazole polyamide oligomers (Gottesfeld *et al.*, 1997, *Nature* 387:202-205; White
et al., 1998, *Nature* 391:468-471), which are specific for the major and minor grooves of a
double helix, respectively.

In addition to synthetic nucleic acids (*i.e.*, antisense, ribozymes, and triplex-forming molecules), there are examples of natural products that interfere with deoxyribonucleic acid ("DNA") or RNA processes such as transcription or translation. For example, certain carbohydrate-based host cell factors, calicheamicin oligosaccharides, interfere with the sequence-specific binding of transcription factors to DNA and inhibit transcription *in vivo* (Ho *et al.*, 1994, Proc. Natl. Acad. Sci. USA 91:9203-9207; Liu *et al.*, 1996, Proc. Natl. Acad. Sci. USA 93:940-944). Certain classes of known antibiotics have been characterized and were found to interact with RNA. For example, the antibiotic thiostreptone binds tightly to a 60-mer from ribosomal RNA (Cundliffe *et al.*, 1990, in *The Ribosome: Structure, Function & Evolution* (Schlessinger *et al.*, eds.) American Society for Microbiology, Washington, D.C. pp. 479-490). Bacterial resistance to various antibiotics often involves methylation at specific rRNA sites (Cundliffe, 1989, Ann. Rev. Microbiol. 43:207-233). Aminoglycosidic aminocyclitol (aminoglycoside) antibiotics and peptide antibiotics are known to inhibit group I intron splicing by binding to specific regions of the RNA (von Ahsen *et al.*, 1991, Nature (London) 353:368-370). Some of these same aminoglycosides have also been found to inhibit hammerhead ribozyme function (Stage *et al.*, 1995, RNA 1:95-101). In addition, certain aminoglycosides and other protein synthesis inhibitors have been found to interact with specific bases in 16S rRNA (Woodcock *et al.*, 1991, EMBO J. 10:3099-3103). An oligonucleotide analog of the 16S rRNA has also been shown to interact with certain aminoglycosides (Purohit *et al.*, 1994, Nature 370:659-662). A molecular basis for hypersensitivity to aminoglycosides has been found to be located in a single base change in mitochondrial rRNA (Hutchin *et al.*, 1993, Nucleic Acids Res. 21:4174-4179). Aminoglycosides have also been shown to inhibit the interaction between specific structural RNA motifs and the corresponding RNA binding protein. Zapp *et al.* (Cell, 1993, 74:969-978) has demonstrated that the aminoglycosides neomycin B, lividomycin A, and tobramycin can block the binding of Rev, a viral regulatory protein required for viral gene expression, to its viral recognition element in the IIB (or RRE) region of HIV RNA. This blockage appears to be the result of competitive binding of the antibiotics directly to the RRE RNA structural motif.

Single stranded sections of RNA can fold into complex tertiary structures consisting of local motifs such as loops, bulges, pseudoknots, guanosine quartets and turns (Chastain & Tinoco, 1991, Progress in Nucleic Acid Res. & Mol. Biol. 41:131-177; Chow & Bogdan, 1997, Chemical Reviews 97:1489-1514; Rando & Hogan, 1998, Biologic activity of guanosine quartet forming oligonucleotides in "Applied Antisense Oligonucleotide Technology" Stein. & Krieg (eds) John Wiley and Sons, New York, pages

335-352). Such structures can be critical to the activity of the nucleic acid and affect functions such as regulation of mRNA transcription, stability, or translation (Weeks & Crothers, 1993; Science 261:1574-1577). The dependence of these functions on the native three-dimensional structural motifs of single-stranded stretches of nucleic acids makes it 5 difficult to identify or design synthetic agents that bind to these motifs using general, simple-to-use sequence-specific recognition rules for the formation of double- and triple-helical nucleic acids used in the design of antisense and ribozyme type molecules.

Approaches to screening generally involve competitive assays designed to identify 10 compounds that disrupt the interaction between a target RNA and a physiological, host cell factor(s) that had been previously identified to specifically interact with that particular target RNA. In general, such assays require the identification and characterization of the host cell factor(s) deemed to be required for the function of the target RNA. Both the target RNA and its preselected host cell binding partner are used in a competitive format to identify 15 compounds that disrupt or interfere with the two components in the assay.

15 Citation or identification of any reference in Section 2 of this application is not an admission that such reference is available as prior art to the present invention.

3. SUMMARY OF THE INVENTION

20 The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids including, but not limited to, specific RNA sequences, RNA structural motifs, and/or RNA structural elements. The specific target RNA sequences, RNA structural motifs, and/or RNA structural elements are used as targets for screening small molecules and identifying those that directly bind these specific 25 sequences, motifs, and/or structural elements. For example, methods are described in which a preselected target RNA having a detectable label is used to screen a library of test compounds, preferably under physiologic conditions. Any complexes formed between the target RNA and a member of the library are identified using physical methods that detect the altered physical property of the target RNA bound to a test compound. In particular, the 30 present invention relates to methods for using a target RNA having a detectable label to screen a library of test compounds free in solution, in labeled tubes or microtiter plate, or in a microarray. Compounds in the library that bind to the labeled target RNA will form a detectably labeled complex. The detectably labeled complex can then be identified and removed from the uncomplexed, unlabeled test compounds in the library, and from 35 uncomplexed, labeled target RNA, by a variety of methods, including but not limited to, methods that differentiate changes in the electrophoretic, chromatographic, or thermostable

properties of the complexed target RNA. Such methods include, but are not limited to, electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation. The structure of the test compound attached to the labeled RNA is then determined. The methods used will depend, in part, on the nature of the library screened. For example, assays or microarrays of test compounds, each having an address or identifier, may be deconvoluted, *e.g.*, by cross-referencing the positive sample to original compound list that was applied to the individual test assays. Another method for identifying test compounds includes *de novo* structure determination of the test compounds using mass spectrometry or nuclear magnetic resonance ("NMR"). The test compounds identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and the like. In addition, small organic molecules which interact specifically with target RNA molecules may be useful as lead compounds for the development of therapeutic agents.

The methods described herein for the identification of compounds that directly bind to a particular preselected target RNA are well suited for high-throughput screening. The direct binding method of the invention offers advantages over drug screening systems for competitors that inhibit the formation of naturally-occurring RNA binding protein:target RNA complexes; *i.e.*, competitive assays. The direct binding method of the invention is rapid and can be set up to be readily performed, *e.g.*, by a technician, making it amenable to high throughput screening. The method of the invention also eliminates the bias inherent in the competitive drug screening systems, which require the use of a preselected host cell factor that may not have physiological relevance to the activity of the target RNA. Instead, the methods of the invention are used to identify any compound that can directly bind to specific target RNA sequences, RNA structural motifs, and/or RNA structural elements, preferably under physiologic conditions. As a result, the compounds so identified can inhibit the interaction of the target RNA with any one or more of the native host cell factors (whether known or unknown) required for activity of the RNA *in vivo*.

The present invention may be understood more fully by reference to the detailed description and examples, which are intended to illustrate non-limiting embodiments of the invention.

3.1. Definitions

As used herein, a "target nucleic acid" refers to RNA, DNA, or a chemically modified variant thereof. In a preferred embodiment, the target nucleic acid is RNA. A target nucleic acid also refers to tertiary structures of the nucleic acids, such as, but not limited to loops, bulges, pseudoknots, guanosine quartets and turns. A target nucleic acid also refers to RNA elements such as, but not limited to, the HIV TAR element, internal ribosome entry site, "slippery site", instability elements, and adenylate uridylate-rich elements, which are described in Section 5.1. Non-limiting examples of target nucleic acids are presented in Section 5.1 and Section 6.

As used herein, a "library" refers to a plurality of test compounds with which a target nucleic acid molecule is contacted. A library can be a combinatorial library, *e.g.*, a collection of test compounds synthesized using combinatorial chemistry techniques, or a collection of unique chemicals of low molecular weight (less than 1000 daltons) that each occupy a unique three-dimensional space.

As used herein, a "label" or "detectable label" is a composition that is detectable, either directly or indirectly, by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive isotopes (*e.g.*, ^{32}P , ^{35}S , and ^3H), dyes, fluorescent dyes, electron-dense reagents, enzymes and their substrates (*e.g.*, as commonly used in enzyme-linked immunoassays, *e.g.*, alkaline phosphatase and horse radish peroxidase), biotin-streptavidin, digoxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available. Moreover, a label or detectable moiety can include a "affinity tag" that, when coupled with the target nucleic acid and incubated with a test compound or compound library, allows for the affinity capture of the target nucleic acid along with molecules bound to the target nucleic acid. One skilled in the art will appreciate that a affinity tag bound to the target nucleic acids has, by definition, a complimentary ligand coupled to a solid support that allows for its capture. For example, useful affinity tags and complimentary partners include, but are not limited to, biotin-streptavidin, complimentary nucleic acid fragments (*e.g.*, oligo dT-oligo dA, oligo T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamers, or haptens and proteins for which antisera or monoclonal antibodies are available. The label or detectable moiety is typically bound, either covalently, through a linker or chemical bound, or through ionic, van der Waals or hydrogen bonds to the molecule to be detected.

As used herein, a "dye" refers to a molecule that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means.

As used herein, a "visible dye" refers to a molecule having a chromophore that absorbs radiation in the visible region of the spectrum (*i.e.*, having a wavelength of between about 400 nm and about 700 nm) such that the transmitted radiation is in the visible region and can be detected either visually or by conventional spectroscopic means. As used herein, an "ultraviolet dye" refers to a molecule having a chromophore that absorbs radiation in the ultraviolet region of the spectrum (*i.e.*, having a wavelength of between about 30 nm and about 400 nm). As used herein, an "infrared dye" refers to a molecule having a chromophore that absorbs radiation in the infrared region of the spectrum (*i.e.*, having a wavelength between about 700 nm and about 3,000 nm). A "chromophore" is the network of atoms of the dye that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means. One of skill in the art will readily appreciate that although a dye absorbs radiation in one region of the spectrum, it may emit radiation in another region of the spectrum. For example, an ultraviolet dye may emit radiation in the visible region of the spectrum. One of skill in the art will also readily appreciate that a dye can transmit radiation or can emit radiation via fluorescence or phosphorescence.

The phrase "pharmaceutically acceptable salt(s)," as used herein includes but is not limited to salts of acidic or basic groups that may be present in test compounds identified using the methods of the present invention. Test compounds that are basic in nature are capable of forming a wide variety of salts with various inorganic and organic acids. The acids that can be used to prepare pharmaceutically acceptable acid addition salts of such basic compounds are those that form non-toxic acid addition salts, *i.e.*, salts containing pharmacologically acceptable anions, including but not limited to sulfuric, citric, maleic, acetic, oxalic, hydrochloride, hydrobromide, hydroiodide, nitrate, sulfate, bisulfate, phosphate, acid phosphate, isonicotinate, acetate, lactate, salicylate, citrate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucaronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate and pamoate (*i.e.*, 1,1'-methylene-bis-(2-hydroxy-3-naphthoate)) salts. Test compounds that include an amino moiety may form pharmaceutically or cosmetically acceptable salts with various amino acids, in addition to the acids mentioned above. Test compounds that are acidic in nature are capable of forming base salts with various pharmacologically or cosmetically acceptable cations. Examples of such salts include alkali metal or alkaline earth metal salts and, particularly, calcium, magnesium, sodium lithium, zinc, potassium, and iron salts.

5 By "substantially one type of test compound," as used herein, is meant that the assay can be performed in such a fashion that at some point, only one compound need be used in each reaction so that, if the result is indicative of a binding event occurring between the target RNA molecule and the test compound, the test compound can be easily identified.

4. DESCRIPTION OF DRAWINGS

FIG. 1. Gel retardation analysis to detect peptide-RNA interactions. In 20 μ l reactions containing increasing concentrations of Tat₄₇₋₅₈ peptide (0.1 μ M, 0.2 μ M, 0.4 μ M, 0.8 μ M, 1.6 μ M) 50 pmole TAR RNA oligonucleotide was added in TK buffer. The reaction mixture was then heated at 90 °C for 2 min and allowed to cool slowly to 24 °C. 10 ml of 30% glycerol was added to each sample and applied to a 12% non-denaturing polyacrylamide gel. The gel was electrophoresed using 1200 volt-hours at 4 °C in TBE Buffer. Following electrophoresis, the gel was dried and the radioactivity was quantitated with a phosphorimager. The concentration of peptide added is indicated above each lane.

15 FIG. 2. Gentamicin interacts with an oligonucleotide corresponding to the 16S rRNA. 20 μ l reactions containing increasing concentrations of gentamicin (1 ng/ml, 10 ng/ml, 100 ng/ml, 1 μ g/ml, 10 μ g/ml, 50 μ g/ml, 500 μ g/ml) were added to 50 pmole RNA oligonucleotide in TKM buffer, heated at 90 °C for 2 min and allowed to cool slowly to 24 °C. Then 10 μ l of 30% glycerol was added to each sample and the samples were applied to a 13.5% non-denaturing polyacrylamide gel. The gel was electrophoresed using 1200 volt-hours at 4 °C in TBE Buffer. Following electrophoresis, the gel was dried and the radioactivity was quantitated using a phosphorimager. The concentration of gentamicin added is indicated above each lane.

20 FIG. 3. The presence of 10 pg/ml gentamicin produces a gel mobility shift in the presence of the 16S rRNA oligonucleotide. 20 μ l reactions containing increasing concentrations of gentamicin (100 ng/ml, 10 ng/ml, 1 ng/ml, 100 pg/ml, and 10 pg/ml) were added to 50 pmole RNA oligonucleotide in TKM buffer were treated as described for Figure 2.

25 FIG. 4. Gentamicin binding to the 16S rRNA oligonucleotide is weak in the absence of MgCl₂. Reaction mixtures containing gentamicin (1 mg/ml, 100 μ g/ml,

30

35

10 $\mu\text{g/ml}$, $1 \mu\text{g/ml}$, $0.1 \mu\text{g/ml}$, and 10 ng/ml) were treated as described in Figure 2 except that the TKM buffer does not contain MgCl_2 .

FIG. 5. Gel retardation analysis to detect peptide-RNA interactions. In reactions 5 containing increasing concentrations of Tat_{47-58} peptide ($0.1 \mu\text{M}$, $0.2 \mu\text{M}$, $0.4 \mu\text{M}$, $0.8 \mu\text{M}$, $1.6 \mu\text{M}$) 50 pmole TAR RNA oligonucleotide was added in TK buffer. The reaction mixture was then heated at 90°C for 2 min and allowed to cool slowly to 24°C . The reactions were loaded onto a SCE9610 automated capillary electrophoresis apparatus (SpectruMedix; State College, Pennsylvania). The peaks correspond to the amount of free TAR RNA ("TAR") or the Tat-TAR complex ("Tat-TAR"). The concentration of peptide added is indicated below each lane.

5. DETAILED DESCRIPTION OF THE INVENTION

15 The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids, in particular, RNAs, including but not limited to preselected target RNA sequencing structural motifs, or structural elements. Methods are described in which a preselected target RNA having a detectable label is used to screen a library of test compounds. Any complexes formed between the target RNA and 20 a member of the library are identified using physical methods that detect the altered physical property of the target RNA bound to a test compound. Changes in the physical property of the RNA-test compound complex relative to the target RNA or test compound can be measured by methods such as, but not limited to, methods that detect a change in mobility due to a change in mass, change in charge, or a change in thermostability. Such methods 25 include, but are not limited to; electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation. In particular, the present invention relates to methods for using a target RNA having a detectable label to screen a library of test compounds free in 30 solution, in labeled tubes or microtiter plate, or in a microarray. Compounds in the library that bind to the labeled target RNA will form a detectably labeled complex. The detectably labeled complex can then be identified and removed from the unlabeled, uncomplexed test compounds in the library by a variety of methods capable of differentiating changes in the physical properties of the complexed target RNA. The structure of the test compound 35 attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. For example, assays or microarrays of test compounds,

each having an address or identifier, may be deconvoluted, e.g., by cross-referencing the positive sample to an original compound list that was applied to the individual test assays. Another method for identifying test compounds includes *de novo* structure determination of the test compounds using mass spectrometry or nuclear magnetic resonance ("NMR").

5 Thus, the methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of test compounds, in which the test compounds of the library that specifically bind a preselected target nucleic acid are easily distinguished from non-binding members of the library. The structures of the binding molecules are deciphered from the input library by methods depending on the type of library that is used.

10 The test compounds so identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and lead compounds for development of therapeutics, and the like. Small organic compounds that are identified to interact specifically with the target RNA molecules are particularly attractive candidates as

15 lead compounds for the development of therapeutic agents.

20 The assay of the invention reduces bias introduced by competitive binding assays which require the identification and use of a host cell factor (presumably essential for modulating RNA function) as a binding partner for the target RNA. The assays of the present invention are designed to detect any compound or agent that binds to the target RNA, preferably under physiologic conditions. Such agents can then be tested for biological activity, without establishing or guessing which host cell factor or factors is required for modulating the function and/or activity of the target RNA.

25 Section 5.1 describes examples of protein-RNA interactions that are important in a variety of cellular functions and several target RNA elements that can be used to identify test compounds. Compounds that inhibit these interactions by binding to the RNA and successfully competing with the natural protein or host cell factor that endogenously binds to the RNA may be important, e.g., in treating or preventing a disease or abnormal condition, such as an infection or unchecked growth. Section 5.2 describes detectable labels for target nucleic acids that are useful in the methods of the invention.

30 Section 5.3 describes libraries of test compounds. Section 5.4 provides conditions for binding a labeled target RNA to a test compound of a library and detecting RNA binding to a test compound using the methods of the invention. Section 5.5 provides methods for separating complexes of target RNAs bound to a test compound from an unbound RNA.

35 Section 5.6 describes methods for identifying test compounds that are bound to the target RNA. Section 5.7 describes a secondary, biological screen of test compounds identified by

the methods of the invention to test the effect of the test compounds *in vivo*. Section 5.8 describes the use of test compounds identified by the methods of the invention for treating or preventing a disease or abnormal condition in mammals.

5

5.1. Biologically Important RNA-Host Cell Factor Interactions

10

Nucleic acids, and in particular RNAs, are capable of folding into complex tertiary structures that include bulges, loops, triple helices and pseudoknots, which can provide binding sites for host cell factors, such as proteins and other RNAs. RNA-protein and RNA-RNA interactions are important in a variety cellular functions, including transcription, RNA splicing, RNA stability and translation. Furthermore, the binding of such host cell factors to RNAs may alter the stability and translational efficiency of such RNAs, and according affect subsequent translation. For example, some diseases are associated with protein overproduction or decreased protein function. In this case, the identification of compounds to modulate RNA stability and translational efficiency will be useful to treat and prevent such diseases.

15

The methods of the present invention are useful for identifying test compounds that bind to target RNA elements in a high throughput screening assay of libraries of test compounds in solution. In particular, the methods of the present invention are useful for identifying a test compound that binds to a target RNA elements and inhibits the interaction of that RNA with one or more host cell factors *in vivo*. The molecules identified using the methods of the invention are useful for inhibiting the formation of a specific bound RNA:host cell factor complexes *in vivo*.

20

In some embodiments, test compounds identified by the methods of the invention are useful for increasing or decreasing the translation of messenger RNAs ("mRNAs"), *e.g.*, protein production, by binding to one or more regulatory elements in the 5' untranslated region, the 3' untranslated region, or the coding region of the mRNA.

25

Compounds that bind to mRNA can, *inter alia*, increase or decrease the rate of mRNA processing, alter its transport through the cell, prevent or enhance binding of the mRNA to ribosomes, suppressor proteins or enhancer proteins, or alter mRNA stability. Accordingly, compounds that increase or decrease mRNA translation can be used to treat or prevent disease. For example, diseases associated with protein overproduction, such as amyloidosis, or with the production of mutant proteins, such as *Ras*, can be treated or prevented by decreasing translation of the mRNA that codes for the overproduced protein, thus inhibiting production of the protein. Conversely, the symptoms of diseases associated with decreased protein function, such as hemophilia, may be treated by increasing

translation of mRNA coding for the protein whose function is decreased, *e.g.*, factor IX in some forms of hemophilia.

The methods of the invention can be used to identify compounds that bind to mRNAs coding for a variety of proteins with which the progression of diseases in mammals is associated. These mRNAs include, but are not limited to, those coding for amyloid protein and amyloid precursor protein; anti-angiogenic proteins such as angiostatin, endostatin, METH-1 and METH-2; apoptosis inhibitor proteins such as survivin, clotting factors such as Factor IX, Factor VIII, and others in the clotting cascade; collagens; cyclins and cyclin inhibitors, such as cyclin dependent kinases, cyclin D1, cyclin E, WAF1, cdk4 inhibitor, and MTS1; cystic fibrosis transmembrane conductance regulator gene (CFTR); cytokines such as IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17 and other interleukins; hematopoietic growth factors such as erythropoietin (Epo); colony stimulating factors such as G-CSF, GM-CSF, M-CSF, SCF and thrombopoietin; growth factors such as BDNF, BMP, GGRP, EGF, FGF, GDNF, GGF, HGF, IGF-1, IGF-2, KGF, myotrophin, NGF, OSM, PDGF, somatotrophin, TGF- β , TGF- α and VEGF; antiviral cytokines such as interferons, antiviral proteins induced by interferons, TNF- α , and TNF- β ; enzymes such as cathepsin K, cytochrome P-450 and other cytochromes, farnesyl transferase, glutathione-S transferases, heparanase, HMG CoA synthetase, N-acetyltransferase, phenylalanine hydroxylase, phosphodiesterase, ras carboxyl-terminal protease, telomerase and TNF converting enzyme; glycoproteins such as cadherins, *e.g.*, N-cadherin and E-cadherin; cell adhesion molecules; selectins; transmembrane glycoproteins such as CD40; heat shock proteins; hormones such as 5- α reductase, atrial natriuretic factor, calcitonin, corticotrophin releasing factor, diuretic hormones, glucagon, gonadotropin, gonadotropin releasing hormone, growth hormone, growth hormone releasing factor, somatotropin, insulin, leptin, luteinizing hormone, luteinizing hormone releasing hormone, parathyroid hormone, thyroid hormone, and thyroid stimulating hormone; proteins involved in immune responses, including antibodies, CTLA4, hemagglutinin, MHC proteins, VLA-4, and kallikrein-kininogen-kinin system; ligands such as CD4; oncogene products such as *sis*, *hst*, protein tyrosine kinase receptors, *ras*, *abl*, *mos*, *myc*, *fos*, *jun*, *H-ras*, *ki-ras*, *c-fms*, *bcl-2*, *L-myc*, *c-myc*, *gip*, *gsp*, and *HER-2*; receptors such as bombesin receptor, estrogen receptor, GABA receptors, growth factor receptors including EGFR, PDGFR, FGFR, and NGFR, GTP-binding regulatory proteins, interleukin receptors, ion channel receptors, leukotriene receptor antagonists, lipoprotein receptors, opioid pain receptors, substance P receptors, retinoic acid and retinoid receptors, steroid receptors, T-cell receptors, thyroid hormone receptors, TNF receptors; tissue

plasminogen activator; transmembrane receptors; transmembrane transporting systems, such as calcium pump, proton pump, Na/Ca exchanger, MRP1, MRP2, P170, LRP, and cMOAT; transferrin; and tumor suppressor gene products such as *APC*, *brca1*, *brca2*, *DCC*, *MCC*, *MTS1*, *NF1*, *NF2*, *nm23*, *p53* and *Rb*. In addition to the eukaryotic genes listed above, the invention, as described, can be used to define molecules that interrupt viral, bacterial or fungal transcription or translation efficiencies and therefore form the basis for a novel anti-infectious disease therapeutic. Other target genes include, but are not limited to, those disclosed in Section 5.1 and Section 6.

The methods of the invention can be used to identify mRNA-binding test compounds for increasing or decreasing the production of a protein, thus treating or preventing a disease associated with decreasing or increasing the production of said protein, respectively. The methods of the invention may be useful for identifying test compounds for treating or preventing a disease in mammals, including cats, dogs, swine, horses, goats, sheep, cattle, primates and humans. Such diseases include, but are not limited to, amyloidosis, hemophilia, Alzheimer's disease, atherosclerosis, cancer, giantism, dwarfism, hypothyroidism, hyperthyroidism, inflammation, cystic fibrosis, autoimmune disorders, diabetes, aging, obesity, neurodegenerative disorders, and Parkinson's disease. Other diseases include, but are not limited to, those described in Section 5.1 and diseases caused by aberrant expression of the genes disclosed in Example 6. In addition to the eukaryotic genes listed above, the invention, as described, can be used to define molecules that interrupt viral, bacterial or fungal transcription or translation efficiencies and therefore form the basis for a novel anti-infectious disease therapeutic.

In other embodiments, test compounds identified by the methods of the invention are useful for preventing the interaction of an RNA, such as a transfer RNA ("tRNA"), an enzymatic RNA or a ribosomal RNA ("rRNA"), with a protein or with another RNA, thus preventing, e.g., assembly of an *in vivo* protein-RNA or RNA-RNA complex that is essential for the viability of a cell. The term "enzymatic RNA," as used herein, refers to RNA molecules that are either self-splicing, or that form an enzyme by virtue of their association with one or more proteins, e.g., as in RNase P, telomerase or small nuclear ribonuclear protein particles. For example, inhibition of an interaction between rRNA and one or more ribosomal proteins may inhibit the assembly of ribosomes, rendering a cell incapable of synthesizing proteins. In addition, inhibition of the interaction of precursor rRNA with ribonucleases or ribonucleoprotein complexes (such as RNase P) that process the precursor rRNA prevent maturation of the rRNA and its assembly into ribosomes. Similarly, a tRNA:tRNA synthetase complex may be inhibited by test

compounds identified by the methods of the invention such that tRNA molecules do not become charged with amino acids. Such interactions include, but are not limited to, rRNA interactions with ribosomal proteins, tRNA interactions with tRNA synthetase, RNase P protein interactions with RNase P RNA, and telomerase protein interactions with telomerase RNA.

In other embodiments, test compounds identified by the methods of the invention are useful for treating or preventing a viral, bacterial, protozoan or fungal infection. For example, transcriptional up-regulation of the genes of human immunodeficiency virus type 1 ("HIV-1") requires binding of the HIV Tat protein to the HIV trans-activation response region RNA ("TAR RNA"). HIV TAR RNA is a 59-base stem-loop structure located at the 5'-end of all nascent HIV-1 transcripts (Jones & Peterlin, 1994, *Annu. Rev. Biochem.* 63:717-43). Tat protein is known to interact with uracil 23 in the bulge region of the stem of TAR RNA. Thus, TAR RNA is a potential binding target for test compounds, such as small peptides and peptide analogs that bind to the bulge region of TAR RNA and inhibit formation of a Tat-TAR RNA complex involved in HIV-1 upregulation (see Hwang *et al.*, 1999 *Proc. Natl. Acad. Sci. USA* 96:12997-13002). Accordingly, test compounds that bind to TAR RNA are useful as anti-HIV therapeutics (Hamy *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:3548-3553; Hamy *et al.*, 1998, *Biochemistry* 37:5086-5095; Mei *et al.*, 1998, *Biochemistry* 37:14204-14212), and therefore, are useful for treating or preventing AIDS.

The methods of the invention can be used to identify test compounds to treat or prevent viral, bacterial, protozoan or fungal infections in a patient. In some embodiments, the methods of the invention are useful for identifying compounds that decrease translation of microbial genes by interacting with mRNA, as described above, or for identifying compounds that inhibit the interactions of microbial RNAs with proteins or other ligands that are essential for viability of the virus or microbe. Examples of microbial target RNAs useful in the present invention for identifying antiviral, antibacterial, anti-protozoan and anti-fungal compounds include, but are not limited to, general antiviral and anti-inflammatory targets such as mRNAs of INF α , INF γ , RNase L, RNase L inhibitor protein, PKR, tumor necrosis factor, interleukins 1-15, and IMP dehydrogenase; internal ribosome entry sites; HIV-1 CT rich domain and RNase H mRNA; HCV internal ribosome entry site (required to direct translation of HCV mRNA), and the 3'-untranslated tail of HCV genomes; rotavirus NSP3 binding site, which binds the protein NSP3 that is required for rotavirus mRNA translation; HBV epsilon domain; Dengue virus 5' and 3' untranslated regions, including IRES; INF α , INF β and INF γ ; plasmodium falciparum mRNAs; the 16S

ribosomal subunit ribosomal RNA and the RNA component of RNase P of bacteria; and the RNA component of telomerase in fungi and cancer cells. Other target viral and bacterial mRNAs include, but are not limited to, those disclosed in Section 6.

5 One of skill in the art will appreciate that, although such target RNAs are functionally conserved in various species (*e.g.*, from yeast to humans), they exhibit nucleotide sequence and structural diversity. Therefore, inhibition of, for example, yeast telomerase by an anti-fungal compound identified by the methods of the invention might not interfere with human telomerase and normal human cell proliferation.

10 Thus, the methods of the invention can be used to identify test compounds that interfere with one or more target RNA interactions with host cell factors that are important for cell growth or viability, or essential in the life cycle of a virus, a bacterium, a protozoa or a fungus. Such test compounds and/or congeners that demonstrate desirable biologic and pharmacologic activity can be administered to a patient in need thereof in order 15 to treat or prevent a disease caused by viral, bacterial, protozoan, or fungal infections. Such diseases include, but are not limited to, HIV infection, AIDS, human T-cell leukemia, SIV infection, FIV infection, feline leukemia, hepatitis A, hepatitis B, hepatitis C, Dengue fever, malaria, rotavirus infection, severe acute gastroenteritis, diarrhea, encephalitis, hemorrhagic fever, syphilis, legionella, whooping cough, gonorrhea, sepsis, influenza, pneumonia, tinea 20 infection, candida infection, and meningitis.

20 Non-limiting examples of RNA elements involved in the regulation of gene expression, *i.e.*, mRNA stability, translational efficiency via translational initiation and 25 ribosome assembly, *etc.*, include the HIV TAR element, internal ribosome entry site, "slippery site", instability elements, and adenylate uridylate-rich elements, as discussed below.

5.1.1. HIV TAR Element

Transcriptional up-regulation of the genes of human immunodeficiency virus type 1 ("HIV-1") requires binding of the HIV Tat protein to the HIV trans-activation 30 response region RNA ("TAR RNA"), a 59-base stem-loop structure located at the 5' end of all nascent HIV-1 transcripts (Jones & Peterlin, 1994, *Annu. Rev. Biochem.* 63:717-43). Tat protein is known to interact with uracil 23 in the bulge region of the stem of TAR RNA. Thus, TAR RNA is a useful binding target for test compounds, such as small peptides and peptide analogs that bind to the bulge region of TAR RNA and inhibit formation of a Tat- 35 TAR RNA complex involved in HIV-1 up-regulation (see Hwang *et al.*, 1999 *Proc. Natl. Acad. Sci. USA* 96:12997-13002). Accordingly, test compounds that bind to TAR RNA

can be useful as anti-HIV therapeutics (Hamy *et al.*, 1997, Proc. Natl. Acad. Sci. USA 94:3548-3553; Hamy *et al.*, 1998, Biochemistry 37:5086-5095; Mei *et al.*, 1998, Biochemistry 37:14204-14212), and therefore, are useful for treating or preventing AIDS.

5

5.1.2. Internal Ribosome Entry Site (“IRES”)

Internal ribosome entry sites (“IRES”) are found in the 5' untranslated regions (“5' UTR”) of several mRNAs, and are thought to be involved in the regulation of translational efficiency. When the IRES element is present on an mRNA downstream of a translational stop codon, it directs ribosomal re-entry (Ghattas *et al.*, 1991, Mol. Cell. Biol. 11:5848-5959), which permits initiation of translation at the start of a second open reading frame.

As reviewed by Jang *et al.*, a large segment of the 5' nontranslated region, approximately 400 nucleotides in length, promotes internal entry of ribosomes independent of the non-capped 5' end of picornavirus mRNAs (mammalian plus-strand RNA viruses whose genomes serve as mRNA). This 400 nucleotide segment (IRES), maps approximately 200 nt down-stream from the 5' end and is highly structured. IRES elements of different picornaviruses, although functionally similar *in vitro* and *in vivo*, are not identical in sequence or structure. However, IRES elements of the genera entero- and rhinoviruses, on the one hand, and cardio- and aphthoviruses, on the other hand, reveal similarities corresponding to phylogenetic kinship. All IRES elements contain a conserved Yn-Xm-AUG unit (Y, pyrimidine; X, nucleotide) which appears essential for IRES function. The IRES elements of cardio-, entero- and aphthoviruses bind a cellular protein, p57. In the case of cardioviruses, the interaction between a specific stem-loop of the IREs is essential for translation *in vitro*. The IRES elements of entero- and cardioviruses also bind the cellular protein, p52, but the significance of this interaction remains to be shown. The function of p57 or p52 in cellular metabolism is unknown. Since picornaviral IRES elements function *in vivo* in the absence of any viral gene products, it is speculated that IRES-like elements may also occur in specific cellular mRNAs releasing them from cap-dependent translation (Jang *et al.*, 1990, Enzyme 44(1-4):292-309).

30

5.1.3. “Slippery Site”

Programmed, or directed, ribosomal frameshifting, when ribosomes shift from one translation reading frame to another and synthesize two viral proteins from a single viral mRNA, is directed by a unique site in viral mRNAs called the “slippery site.” The slippery site directs ribosomal frameshifting in the -1 or +1 direction that causes the

ribosome to slip by one base in the 5' direction thereby placing the ribosome in the new reading frame to produce a new protein.

Programmed, or directed, ribosomal frameshifting is of particular value to viruses that package their plus strands, as it eliminates the need to splice their mRNAs and reduces the risk of packaging defective genomes and regulates the ratio of viral proteins synthesized. Examples of programmed translational frameshifting (both +1 and -1 shifts) have been identified in ScV systems (Lopinski *et al.*, 2000, Mol. Cell. Biol. 20(4):1095-103, retroviruses (Falk *et al.*, 1993, J. Virol. 67:273-6277; Jacks & Varmus, 1985, Science 230:1237-1242; Morikawa & Bishop, 1992, Virology 186:389-397; Nam *et al.*, 1993, J. Virol. 67:196-203); coronaviruses (Brierley *et al.*, 1987, EMBO J. 6:3779-3785; Herold & Siddell, 1993, Nucleic Acids Res. 21:5838-5842); giardiviruses, which are also members of the Totiviridae (Wang *et al.*, 1993, Proc. Natl. Acad. Sci. USA 90:8595-8599); two bacterial genes (Blinkowa & Walker, 1990, Nucleic Acids Res., 18:1725-1729; Craigen & Caskey, 1986, Nature 322:273); bacteriophage genes (Condron *et al.*, 1991, Nucleic Acids Res. 19:5607-5612); astroviruses (Marczinke *et al.*, 1994, J. Virol. 68:5588-5595); the yeast EST3 gene (Lundblad & Morris, 1997, Curr. Biol. 7:969-976); and the rat, mouse, Xenopus, and Drosophila ornithine decarboxylase antizymes (Matsufuji *et al.*, 1995, Cell 80:51-60); and a significant number of cellular genes (Herold & Siddell, 1993, Nucleic Acids Res. 21:5838-5842).

Drugs targeted to ribosomal frameshifting minimize the problem of virus drug resistance because this strategy targets a host cellular process rather than one introduced into the cell by the virus, which minimizes the ability of viruses to evolve drug-resistant mutants. Compounds that target the RNA elements involved in regulating programmed frameshifting should have several advantages, including (a) any selective pressure on the host cellular translational machinery to adapt to the drugs would have to occur at the host evolutionary time scale, which is on the order of millions of years, (b) ribosomal frameshifting is not used to express any host proteins, and (c) altering viral frameshifting efficiencies by modulating the activity of a host protein minimizing the likelihood that the virus will acquire resistance to such inhibition by mutations in its own genome.

5.1.4. Instability Elements

“Instability elements” may be defined as specific sequence elements that promote the recognition of unstable mRNAs by cellular turnover machinery. Instability

elements have been found within mRNA protein coding regions as well as untranslated regions.

5 Altering the control of stability of normal mRNAs may lead to disease. The alteration of mRNA stability has been implicated in diseases such as, but not limited to, cancer, immune disorders, heart disease, and fibrotic disorders.

There are several examples of mutations that delete instability elements which then result in stabilization of mRNAs that may be involved in the onset of cancer. In Burkitt's lymphoma, a portion of the *c-myc* proto-oncogene is translocated to an Ig locus, producing a form of the *c-myc* mRNA that is five times more stable (see, e.g., Kapstein *et al.*, 1996, *J. Biol. Chem.* 271(31):18875-84). The highly oncogenic *v-fos* mRNA lacks the 10 3' UTR adenylate uridylate rich element ("ARE") that is found in the more labile and weakly oncogenic *c-fos* mRNA (see, e.g., Schiavi *et al.*, 1992, *Biochim Biophys Acta*, 1114(2-3):95-106). Differences between the benign cervical lesions brought about by 15 nonintegrated circular human papillomavirus type 16 and its integrated form, that lacks the 3' UTR ARE and correlates with cervical carcinomas, may be a consequence of stabilizing the E6/E7 transcripts encoding oncogenic proteins. Integration of the virus results in deletion of the ARE instability element, resulting in stabilization of the transcripts and over-expression of the proteins (see, e.g., Jeon & Lambert, 1995, *Proc. Natl. Acad. Sci. USA* 92(5):1654-8). Deletion of AREs from the 3' UTR of the IL-2 and IL-3 genes promotes 20 increased stabilization of these mRNAs, high expression of these proteins, and leads to the formation of cancerous cells (see, e.g., Stoecklin *et al.*, 2000, *Mol. Cell. Biol.* 20(11):3753-63).

Mutations in trans-acting factors involved in mRNA turnover may also 25 promote cancer. In monocytic tumors, the lymphokine GM-CSF mRNA is specifically stabilized as a consequence of an oncogenic lesion in a trans-acting factor that controls mRNA turnover rates. Furthermore, the normally unstable IL-3 transcript is inappropriately long-lived in mast tumor cells. Similarly, the labile GM-CSF mRNA is greatly stabilized in bladder carcinoma cells. See, e.g., Bickel *et al.*, 1990, *J. Immunol.* 145(3):840-5.

30 The immune system is regulated by a large number of regulatory molecules that either activate or inhibit the immune response. It has now been clearly demonstrated that stability of the transcripts encoding these proteins are highly regulated. Altered regulation of these molecules leads to mis-regulation of this process and can result in drastic medical consequences. For example, recent results using transgenic mice have shown that 35 mis-regulation of the stability of the important modulator TNF α mRNA leads to diseases

such as, but not limited to, rheumatoid arthritis and a Crohn's-like liver disease. *See, e.g.*, Clark, 2000, *Arthritis Res.* 2(3):172-4.

Smooth muscle in the heart is modulated by the β -adrenergic receptor, which 5 in turn responds to the sympathetic neurotransmitter norepinephrine and the adrenal hormone epinephrine. Chronic heart failure is characterized by impairment of smooth muscle cells, which results, in part, from the more rapid decay of the β -adrenergic receptor mRNA. *See, e.g.*, Ellis & Frielle, 1999, *Biochem. Biophys. Res. Commun.* 258(3):552-8.

A large number of diseases result from over-expression of collagen. For 10 example, cirrhosis results from damage to the liver as a consequence of cancer, viral infection, or alcohol abuse. Such damage causes mis-regulation of collagen expression, leading to the formation of large collagen deposits. Recent results indicate that the sizeable increase in collagen expression is largely attributable to stabilization of its mRNA. *See, e.g.*, Lindquist *et al.*, 2000, *Am. J. Physiol. Gastrointest. Liver Physiol.* 279(3):G471-6.

15

5.1.5. Adenylate Uridylate-rich Elements (“ARE”)

Adenylate uridylate-rich elements (“ARE”) are found in the 3' untranslated regions (“3' UTR”) of several mRNAs, and involved in the turnover of mRNAs, such as but not limited to transcription factors, cytokines, and lymphokines. AREs may function both 20 as stabilizing and destabilizing elements. ARE mRNAs are classified into five groups, depending on sequence (Bakheet *et al.*, 2001, *Nucl. Acids Res.* 29(1):246-254). An ongoing database at the web site <http://rc.kfshrc.edu.sa/ared> contains ARE-containing mRNAs and their cluster groups, which is incorporated by reference in its entirety. The ARE motifs are classified as follows:

25	Group I Cluster	(AUUUAUUUAUUUAUUUAUUU)	SEQ ID NO: 1
	Group II Cluster	(AUUUAUUUAUUUAUUU) stretch	SEQ ID NO: 2
	Group III Cluster	(WAUUUAUUUAUUUAW) stretch	SEQ ID NO: 3
	Group IV Cluster	(WWAUUUAUUUAWW) stretch	SEQ ID NO: 4
30	Group V Cluster	(WWWWAUUUAWWWWW) stretch	SEQ ID NO: 5

The ARE-mRNAs were clustered into five groups containing five, four, three and two pentameric repeats, while the last group contains only one pentamer within the 13-bp ARE pattern. Functional categories were assigned whenever possible according to NCBI-COG functional annotation (Tatusov *et al.*, 2001, *Nucleic Acids Research*, 29(1): 35 22-28), in addition to the categories: inflammation, immune response, development/differentiation, using an extensive literature search.

5 Group I contains many secreted proteins including GM-CSF, IL-1, IL-11, IL-12 and Gro- β that affect the growth of hematopoietic and immune cells (Witsell & Schook, 1992, Proc. Natl Acad. Sci. USA, 89:4754–4758). Although TNF α is both a pro-inflammatory and anti-tumor protein, there is experimental evidence that it can act as a growth factor in certain leukemias and lymphomas (Liu *et al.*, 2000, J. Biol. Chem. 275:21086–21093).

10 Unlike Group I, Groups II–V contain functionally diverse gene families comprising immune response, cell cycle and proliferation, inflammation and coagulation, angiogenesis, metabolism, energy, DNA binding and transcription, nutrient transportation and ionic homeostasis, protein synthesis, cellular biogenesis, signal transduction, and apoptosis (Bakheet *et al.*, 2001, Nucl. Acids Res. 29(1):246-254).

15 Several groups have described ARE-binding proteins that influence the ARE-mRNA stability. Among the well-characterized proteins are the mammalian homologs of ELAV (embryonic lethal abnormal vision) proteins including AUF1, HuR and He1-N2 (Zhang *et al.*, 1993, Mol. Cell. Biol. 13:7652–7665; Levine *et al.*, 1993, Mol. Cell. Biol. 13:3494–3504; Ma *et al.*, 1996, J. Biol. Chem. 271:8144–8151). The zinc-finger protein tristetraprolin has been identified as another ARE-binding protein with destabilizing activity on TNF α , IL-3 and GM-CSF mRNAs (Stoecklin *et al.*, 2000, Mol. Cell. Biol. 20:3753–3763; Carballo *et al.*, 2000, Blood 95:1891–1899).

20 Since ARE-containing genes are clearly important in biological systems, including but not limited to a number of the early response genes that regulate cell proliferation and responses to exogenous agents, the identification of compounds that bind to one or more of the ARE clusters and potentially modulate the stability of the target RNA can potentially be of value as a therapeutic.

5.2. Detectably Labeled Target RNAs

25 Target nucleic acids, including but not limited to RNA and DNA, useful in the methods of the present invention have a label that is detectable via conventional spectroscopic means or radiographic means. Preferably, target nucleic acids are labeled with a covalently attached dye molecule. Useful dye-molecule labels include, but are not limited to, fluorescent dyes, phosphorescent dyes, ultraviolet dyes, infrared dyes, and visible dyes. Preferably, the dye is a visible dye.

30 Useful labels in the present invention can include, but are not limited to, spectroscopic labels such as fluorescent dyes (*e.g.*, fluorescein and derivatives such as fluorescein isothiocyanate (FITC) and Oregon Green™, rhodamine and derivatives (*e.g.*,

Texas red, tetramethylrhodamine isothiocyanate (TRITC), bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, *etc.*), digoxigenin, biotin, phycoerythrin, AMCA, CyDye™, and the like), radiolabels (*e.g.*, ^3H , ^{125}I , ^{35}S , ^{14}C , ^{32}P , ^{33}P , *etc.*), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase *etc.*), spectroscopic colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.* polystyrene, polypropylene, latex, *etc.*) beads, or nanoparticles – nanoclusters of inorganic ions with defined dimension from 0.1 to 1000 nm. Useful affinity tags and complimentary partners include, but are not limited to, biotin-streptavidin, complimentary nucleic acid fragments (*e.g.*, oligo dT-oligo dA, oligo 5 T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamer-streptavidin, or haptens and 10 proteins for which antisera or monoclonal antibodies are available. The label may be coupled directly or indirectly to a component of the detection assay (*e.g.*, the detection reagent) according to methods well known in the art. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the 15 compound, stability requirements, available instrumentation, and disposal provisions.

15 In one embodiment, nucleic acids that are labeled at one or more specific locations are chemically synthesized using phosphoramidite or other solution or solid-phase methods. Detailed descriptions of the chemistry used to form polynucleotides by the phosphoramidite method are well known (*see, e.g.*, Caruthers *et al.*, U.S. Pat. Nos. 20 4,458,066 and 4,415,732; Caruthers *et al.*, 1982, *Genetic Engineering* 4:1-17; *Users Manual Model 392 and 394 Polynucleotide Synthesizers*, 1990, pages 6-1 through 6-22, Applied Biosystems, Part No. 901237; Ojwang, *et al.*, 1997, *Biochemistry*, 36:6033-6045). The phosphoramidite method of polynucleotide synthesis is the preferred method because of its efficient and rapid coupling and the stability of the starting materials. The synthesis is 25 performed with the growing polynucleotide chain attached to a solid support, such that excess reagents, which are generally in the liquid phase, can be easily removed by washing, decanting, and/or filtration, thereby eliminating the need for purification steps between synthesis cycles.

30 The following briefly describes illustrative steps of a typical polynucleotide synthesis cycle using the phosphoramidite method. First, a solid support to which is attached a protected nucleoside monomer at its 3' terminus is treated with acid, *e.g.*, trichloroacetic acid, to remove the 5'-hydroxyl protecting group, freeing the hydroxyl group for a subsequent coupling reaction. After the coupling reaction is completed an activated intermediate is formed by contacting the support-bound nucleoside with a protected 35 nucleoside phosphoramidite monomer and a weak acid, *e.g.*, tetrazole. The weak acid protonates the nitrogen atom of the phosphoramidite forming a reactive intermediate.

Nucleoside addition is generally complete within 30 seconds. Next, a capping step is performed, which terminates any polynucleotide chains that did not undergo nucleoside addition. Capping is preferably performed using acetic anhydride and 1-methylimidazole. 5 The phosphite group of the internucleotide linkage is then converted to the more stable phosphotriester by oxidation using iodine as the preferred oxidizing agent and water as the oxygen donor. After oxidation, the hydroxyl protecting group of the newly added nucleoside is removed with a protic acid, *e.g.*, trichloroacetic acid or dichloroacetic acid, and the cycle is repeated one or more times until chain elongation is complete. After 10 synthesis, the polynucleotide chain is cleaved from the support using a base, *e.g.*, ammonium hydroxide or *t*-butyl amine. The cleavage reaction also removes any phosphate protecting groups, *e.g.*, cyanoethyl. Finally, the protecting groups on the exocyclic amines of the bases and any protecting groups on the dyes are removed by treating the 15 polynucleotide solution in base at an elevated temperature, *e.g.*, at about 55°C. Preferably the various protecting groups are removed using ammonium hydroxide or *t*-butyl amine.

Any of the nucleoside phosphoramidite monomers can be labeled using standard phosphoramidite chemistry methods (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002; Ojwang *et al.*, 1997, Biochemistry. 36:6033-6045 and references cited therein). Dye molecules useful for covalently coupling to phosphoramidites preferably 20 comprise a primary hydroxyl group that is not part of the dye's chromophore. Illustrative dye molecules include, but are not limited to, disperse dye CAS 4439-31-0, disperse dye CAS 6054-58-6, disperse dye CAS 4392-69-2 (Sigma-Aldrich, St. Louis, MO), disperse red, and 1-pyrenebutanol (Molecular Probes, Eugene, OR). Other dyes useful for coupling 25 to phosphoramidites will be apparent to those of skill in the art, such as fluorescein, cy3, and cy5 fluorescent dyes, and may be purchased from, *e.g.*, Sigma-Aldrich, St. Louis, MO or Molecular Probes, Inc., Eugene, OR.

In another embodiment, dye-labeled target RNA molecules are synthesized enzymatically using *in vitro* transcription (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). In this embodiment, a template DNA is 30 denatured by heating to about 90°C and an oligonucleotide primer is annealed to the template DNA, for example by slow-cooling the mixture of the denatured template and the primer from about 90°C to room temperature. A mixture of ribonucleoside-5'-triphosphates capable of supporting template-directed enzymatic extension of the primed template (*e.g.*, a mixture including GTP, ATP, CTP, and UTP), including one or more dye-labeled 35 ribonucleotides (Sigma-Aldrich, St. Louis, MO), is added to the primed template. Next, a polymerase enzyme is added to the mixture under conditions where the polymerase enzyme

is active, which are well-known to those skilled in the art. A labeled polynucleotide is formed by the incorporation of the labeled ribonucleotides during polymerase-mediated strand synthesis.

5 In yet another embodiment of the invention, nucleic acid molecules are end-labeled after their synthesis. Methods for labeling the 5'-end of an oligonucleotide include but are by no means limited to: (i) periodate oxidation of a 5'-to-5'-coupled ribonucleotide, followed by reaction with an amine-reactive label (Heller & Morisson, 1985, in *Rapid Detection and Identification of Infectious Agents*, D.T. Kingsbury and S. Falkow, eds., pp. 245-256, Academic Press); (ii) condensation of ethylenediamine with 5'-phosphorylated 10 polynucleotide, followed by reaction with an amine reactive label (Morrison, European Patent Application 232 967); (iii) introduction of an aliphatic amine substituent using an aminohexyl phosphite reagent in solid-phase DNA synthesis, followed by reaction with an amine reactive label (Cardullo *et al.*, 1988, Proc. Natl. Acad. Sci. USA 85:8790-8794); and 15 (iv) introduction of a thiophosphate group on the 5'-end of the nucleic acid, using phosphatase treatment followed by end-labeling with ATP-³²S and kinase, which reacts specifically and efficiently with maleimide-labeled fluorescent dyes (Czworkowski *et al.*, 1991, Biochem. 30:4821-4830).

20 A detectable label should not be incorporated into a target nucleic acid at the specific binding site at which test compounds are likely to bind, since the presence of a 25 covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site. Accordingly, if the region of the target nucleic acid that binds to a host cell factor is known, a detectable label is preferably incorporated into the nucleic acid molecule at one or more positions that are spatially or sequentially remote from the binding region.

25 After synthesis, the labeled target nucleic acid can be purified using standard techniques known to those skilled in the art (see Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). Depending on the length of the target nucleic acid and the method of its synthesis, such purification techniques include, but 30 are not limited to, reverse-phase high-performance liquid chromatography ("reverse-phase HPLC"), fast performance liquid chromatography ("FPLC"), and gel purification. After purification, the target RNA is refolded into its native conformation, preferably by heating to approximately 85-95°C and slowly cooling to room temperature in a buffer, e.g., a buffer comprising about 50 mM Tris-HCl, pH 8 and 100 mM NaCl.

35 In another embodiment, the target nucleic acid can also be radiolabeled. A radiolabel, such as, but not limited to, an isotope of phosphorus, sulfur, or hydrogen, may be

incorporated into a nucleotide, which is added either after or during the synthesis of the target nucleic acid. Methods for the synthesis and purification of radiolabeled nucleic acids are well known to one of skill in the art. See, e.g., Sambrook *et al.*, 1989, in *Molecular Cloning: A Laboratory Manual*, pp 10.2-10.70, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entireties.

In another embodiment, the target nucleic acid can be attached to an inorganic nanoparticle. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to 10 Ag_2S , ZnS , CdS , CdTe , Au , or TiO_2 . Nanoparticles have unique optical, electronic and catalytic properties relative to bulk materials which can be adjusted according to the size of the particle. Methods for the attachment of nucleic acids are well known to one of skill in the art (see, e.g., Niemeyer, 2001, *Angew. Chem. Int. Ed.* 40: 4129-4158, International Patent Publication WO/0218643, and the references cited therein, the disclosures of which are hereby incorporated by reference in their entireties).

5.3. Libraries of Small Molecules

Libraries screened using the methods of the present invention can comprise a variety of types of test compounds. In some embodiments, the test compounds are nucleic acid or peptide molecules. In a non-limiting example, peptide molecules can exist in a 20 phage display library. In other embodiments, types of test compounds include, but are not limited to, peptide analogs including peptides comprising non-naturally occurring amino acids, e.g., D-amino acids, phosphorous analogs of amino acids, such as α -amino phosphoric acids and α -amino phosphoric acids, or amino acids having non-peptide 25 linkages, nucleic acid analogs such as phosphorothioates and PNAs, hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose. Libraries of polypeptides or proteins can also be used.

In a preferred embodiment, the combinatorial libraries are small organic 30 molecule libraries, such as, but not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, and diazepindiones. In another embodiment, the combinatorial libraries comprise peptoids; random bio-oligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; 35 vinyllogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; or carbohydrate libraries.

Combinatorial libraries are themselves commercially available (see, e.g., Advanced ChemTech Europe Ltd., Cambridgeshire, UK; ASINEX, Moscow Russia; BioFocus plc, Sittingbourne, UK; Bionet Research (A division of Key Organics Limited), Camelford, UK; ChemBridge Corporation, San Diego, California; ChemDiv Inc, San Diego, California.; ChemRx Advanced Technologies, South San Francisco, California; ComGenex Inc., Budapest, Hungary; Evotec OAI Ltd, Abingdon, UK; IF LAB Ltd., Kiev, Ukraine; Maybridge plc, Cornwall, UK; PharmaCore, Inc., North Carolina; SIDDCO Inc, Tucson, Arizona; TimTec Inc, Newark, Delaware; Tripos Receptor Research Ltd, Bude, UK; Toslab, Ekaterinburg, Russia).

10 In one embodiment, the combinatorial compound library for the methods of the present invention may be synthesized. There is a great interest in synthetic methods directed toward the creation of large collections of small organic compounds, or libraries, which could be screened for pharmacological, biological or other activity (Dolle, 2001, J. Comb. Chem. 3:477-517; Hall *et al.*, 2001, J. Comb. Chem. 3:125-150; Dolle, 2000, J. Comb. Chem. 2:383-433; Dolle, 1999, J. Comb. Chem. 1:235-282). The synthetic methods applied to create vast combinatorial libraries are performed in solution or in the solid phase, *i.e.*, on a solid support. Solid-phase synthesis makes it easier to conduct multi-step reactions and to drive reactions to completion with high yields because excess reagents can be easily added and washed away after each reaction step. Solid-phase combinatorial synthesis also tends to improve isolation, purification and screening. However, the more traditional solution phase chemistry supports a wider variety of organic reactions than solid-phase chemistry. Methods and strategies for the synthesis of combinatorial libraries can be found in *A Practical Guide to Combinatorial Chemistry*, A.W. Czarnik and S.H. Dewitt, eds., American Chemical Society, 1997; *The Combinatorial Index*, B.A. Bunin, Academic Press, 1998; *Organic Synthesis on Solid Phase*, F.Z. Dörwald, Wiley-VCH, 2000; and *Solid-Phase Organic Syntheses*, Vol. 1, A.W. Czarnik, ed., Wiley Interscience, 2001.

20 Combinatorial compound libraries of the present invention may be synthesized using apparatuses described in US Patent No. 6,358,479 to Frisina *et al.*, U.S. Patent No. 6,190,619 to Kilcoin *et al.*, US Patent No. 6,132,686 to Gallup *et al.*, US Patent No. 6,126,904 to Zuellig *et al.*, US Patent No. 6,074,613 to Harness *et al.*, US Patent No. 6,054,100 to Stanchfield *et al.*, and US Patent No. 5,746,982 to Saneii *et al.* which are hereby incorporated by reference in their entirety. These patents describe synthesis apparatuses capable of holding a plurality of reaction vessels for parallel synthesis of 25 multiple discrete compounds or for combinatorial libraries of compounds.

In one embodiment, the combinatorial compound library can be synthesized in solution. The method disclosed in U.S. Patent No. 6,194,612 to Boger *et al.*, which is hereby incorporated by reference in its entirety, features compounds useful as templates for solution phase synthesis of combinatorial libraries. The template is designed to permit reaction products to be easily purified from unreacted reactants using liquid/liquid or solid/liquid extractions. The compounds produced by combinatorial synthesis using the template will preferably be small organic molecules. Some compounds in the library may mimic the effects of non-peptides or peptides. In contrast to solid phase synthesis of combinatorial compound libraries, liquid phase synthesis does not require the use of 5 specialized protocols for monitoring the individual steps of a multistep solid phase synthesis (Egner *et al.*, 1995, *J. Org. Chem.* 60:2652; Anderson *et al.*, 1995, *J. Org. Chem.* 60:2650; Fitch *et al.*, 1994, *J. Org. Chem.* 59:7955; Look *et al.*, 1994, *J. Org. Chem.* 49:7588; 10 Metzger *et al.*, 1993, *Angew. Chem., Int. Ed. Engl.* 32:894; Youngquist *et al.*, 1994, *Rapid Commun. Mass Spect.* 8:77; Chu *et al.*, 1995, *J. Am. Chem. Soc.* 117:5419; Brummel *et al.*, 15 1994, *Science* 264:399; Stevanovic *et al.*, 1993, *Bioorg. Med. Chem. Lett.* 3:431).

Combinatorial compound libraries useful for the methods of the present invention can be synthesized on solid supports. In one embodiment, a split synthesis method, a protocol of separating and mixing solid supports during the synthesis, is used to 20 synthesize a library of compounds on solid supports (see Lam *et al.*, 1997, *Chem. Rev.* 97:41-448; Ohlmeyer *et al.*, 1993, *Proc. Natl. Acad. Sci. USA* 90:10922-10926 and references cited therein). Each solid support in the final library has substantially one type of 25 test compound attached to its surface. Other methods for synthesizing combinatorial libraries on solid supports, wherein one product is attached to each support, will be known to those of skill in the art (see, e.g., Nefzi *et al.*, 1997, *Chem. Rev.* 97:449-472 and US Patent No. 6,087,186 to Cargill *et al.* which are hereby incorporated by reference in their 30 entirety).

As used herein, the term "solid support" is not limited to a specific type of solid support. Rather a large number of supports are available and are known to one skilled in the art. Solid supports include silica gels, resins, derivatized plastic films, glass beads, 35 cotton, plastic beads, polystyrene beads, alumina gels, and polysaccharides. A suitable solid support may be selected on the basis of desired end use and suitability for various synthetic protocols. For example, for peptide synthesis, a solid support can be a resin such as p-methylbenzhydrylamine (pMBHA) resin (Peptides International, Louisville, KY), polystyrenes (e.g., PAM-resin obtained from Bachem Inc., Peninsula Laboratories, etc.), 35 including chloromethylpolystyrene, hydroxymethylpolystyrene and

aminomethylpolystyrene, poly (dimethylacrylamide)-grafted styrene co-divinyl-benzene (e.g., POLYHIPE resin, obtained from Aminotech, Canada), polyamide resin (obtained from Peninsula Laboratories), polystyrene resin grafted with polyethylene glycol (e.g., TENTAGEL or ARGOGEL, Bayer, Tubingen, Germany) polydimethylacrylamide resin (obtained from Milligen/Bioscience, California), or Sepharose (Pharmacia, Sweden).

5 In one embodiment, the solid phase support is suitable for *in vivo* use, *i.e.*, it can serve as a carrier or support for administration of the test compound to a patient (e.g., TENTAGEL, Bayer, Tubingen, Germany). In a particular embodiment, the solid support is 10 palatable and/or orally ingestable.

10 In some embodiments of the present invention, compounds can be attached to solid supports via linkers. Linkers can be integral and part of the solid support, or they may be nonintegral that are either synthesized on the solid support or attached thereto after 15 synthesis. Linkers are useful not only for providing points of test compound attachment to the solid support, but also for allowing different groups of molecules to be cleaved from the solid support under different conditions, depending on the nature of the linker. For example, linkers can be, *inter alia*, electrophilically cleaved, nucleophilically cleaved, photocleavable, enzymatically cleaved, cleaved by metals, cleaved under reductive 20 conditions or cleaved under oxidative conditions.

20 In another embodiment, the combinatorial compound libraries can be assembled *in situ* using dynamic combinatorial chemistry as described in European Patent Application 1,118,359 A1 to Lehn; Huc & Nguyen, 2001, Comb. Chem. High Throughput. Screen. 4:53-74; Lehn and Eliseev, 2001, Science 291:2331-2332; Cousins *et al.* 2000, Curr. Opin. Chem. Biol. 4: 270-279; and Karan & Miller, 2000, Drug. Disc. Today 5:67-75 which are incorporated by reference in their entirety.

25 Dynamic combinatorial chemistry uses non-covalent interaction with a target biomolecule, including but not limited to a protein, RNA, or DNA, to favor assembly of the most tightly binding molecule that is a combination of constituent subunits present as a mixture in the presence of the biomolecule. According to the laws of thermodynamics, 30 when a collection of molecules is able to combine and recombine at equilibrium through reversible chemical reactions in solution, molecules, preferably one molecule, that bind most tightly to a templating biomolecule will be present in greater amount than all other possible combinations. The reversible chemical reactions include, but are not limited to, imine, acyl-hydrazone, amide, acetal, or ester formation between carbonyl-containing 35 compounds and amines, hydrazines, or alcohols; thiol exchange between disulfides; alcohol

exchange in borate esters; Diels-Alder reactions; thermal- or photoinduced sigmatropic or electrocyclic rearrangements; or Michael reactions.

In the preferred embodiment of this technique, the constituent components of the dynamic combinatorial compound library are allowed to combine and reach equilibrium in the absence of the target RNA and then incubated in the presence of the target RNA, preferably at physiological conditions, until a second equilibrium is reached. The second, perturbed, equilibrium (the so-called "templated mixture") can, but need not necessarily, be fixed by a further chemical transformation, including but not limited to reduction, oxidation, hydrolysis, acidification, or basification, to prevent restoration of the original equilibrium when the dynamical combinatorial compound library is separated from the target RNA.

In the preferred embodiment of this technique, the predominant product or products of the templated dynamic combinatorial library can be separated from the minor products and directly identified. In another embodiment, the identity of the predominant product or products can be identified by a deconvolution strategy involving preparation of derivative dynamic combinatorial libraries, as described in European Patent Application 1,118,359 A1, which is incorporated by reference in their entirety, whereby each component of the mixture is, preferably one-by-one but possibly group-wise, left out of the mixture and the ability of the derivative library mixture at chemical equilibrium to bind the target RNA is measured. The components whose removal most greatly reduces the ability of the derivative dynamic combinatorial library to bind the target RNA are likely the components of the predominant product or products in the original dynamic combinatorial library.

25

5.4. Library Screening

After a target nucleic acid, such as but not limited to RNA or DNA, is labeled and a test compound library is synthesized or purchased or both, the labeled target nucleic acid is used to screen the library to identify test compounds that bind to the nucleic acid. Screening comprises contacting a labeled target nucleic acid with an individual, or small group, of the components of the compound library. Preferably, the contacting occurs in an aqueous solution, and most preferably, under physiologic conditions. The aqueous solution preferably stabilizes the labeled target nucleic acid and prevents denaturation or degradation of the nucleic acid without interfering with binding of the test compounds. The aqueous solution can be similar to the solution in which a complex between the target RNA and its corresponding host cell factor (if known) is formed *in vitro*. For example, TK

buffer, which is commonly used to form Tat protein-TAR RNA complexes *in vitro*, can be used in the methods of the invention as an aqueous solution to screen a library of test compounds for TAR RNA binding compounds.

5 The methods of the present invention for screening a library of test compounds preferably comprise contacting a test compound with a target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions. The aqueous solution optionally further comprises non-specific nucleic acids, such as, but not limited to, DNA; yeast tRNA; salmon sperm DNA; homoribopolymers such as, but not 10 limited to, poly IC, polyA, polyU, and polyC; and non-specific RNA. The non-specific RNA may be an unlabeled target nucleic acid having a mutation at the binding site, which renders the unlabeled nucleic acid incapable of interacting with a test compound at that site. For example, if dye-labeled TAR RNA is used to screen a library, unlabeled TAR RNA 15 having a mutation in the uracil 23/cytosine 24 bulge region may also be present in the aqueous solution. Without being bound by any theory, the addition of unlabeled RNA that is essentially identical to the dye-labeled target RNA except for a mutation at the binding site might minimize interactions of other regions of the dye-labeled target RNA with test compounds or with the solid support and prevent false positive results.

20 The solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. The pH of the solution typically ranges from about 5 to about 8, preferably from about 6 to about 8, most preferably from about 6.5 to about 8. A variety of buffers may be used to achieve the desired pH. Suitable buffers include, but are not limited to, Tris, Mes, Bis-Tris, Ada, Aces, Pipes, Mopso, Bis-Tris propane, Bes, 25 Mops, Tes, Hepes, Dipso, Mobs, Tapso, Trizma, Heppso, Popso, TEA, Epps, Tricine, Gly-Gly, Bicine, and sodium-potassium phosphate. The buffering agent comprises from about 10 mM to about 100 mM, preferably from about 25 mM to about 75 mM, most preferably from about 40 mM to about 60 mM buffering agent. The pH of the aqueous solution can be optimized for different screening reactions, depending on the target RNA used and the 30 types of test compounds in the library, and therefore, the type and amount of the buffer used in the solution can vary from screen to screen. In a preferred embodiment, the aqueous solution has a pH of about 7.4, which can be achieved using about 50 mM Tris buffer.

35 In addition to an appropriate buffer, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl₂. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂. Without

being bound by any theory, Applicant has found that a combination of KCl, NaCl, and MgCl₂ stabilizes the target RNA such that most of the RNA is not denatured or digested over the course of the screening reaction. The optional concentration of each salt used in the aqueous solution is dependent on the particular target RNA used and can be determined using routine experimentation.

The solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant. Without being bound by any theory, a small amount of detergent or surfactant in the solution might reduce non-specific binding of the target RNA to the solid support and control aggregation and increase stability of target RNA molecules.

Typical detergents useful in the methods of the present invention include, but are not limited to, anionic detergents, such as salts of deoxycholic acid, 1-heptanesulfonic acid, N-laurylsarcosine, lauryl sulfate, 1-octane sulfonic acid and taurocholic acid; cationic detergents such as benzalkonium chloride, cetylpyridinium, methylbenzethonium chloride, and decamethonium bromide; zwitterionic detergents such as CHAPS, CHAPSO, alkyl betaines, alkyl amidoalkyl betaines, N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, and phosphatidylcholine; and non-ionic detergents such as n-decyl α -D-glucopyranoside, n-decyl β -D-maltopyranoside, n-dodecyl β -D-maltoside, n-octyl β -D-glucopyranoside, sorbitan esters, n-tetradecyl β -D-maltoside, octylphenoxy polyethoxyethanol (Nonidet P-40), nonylphenoxy polyethoxyethanol (NP-40), and tritons.

Preferably, the detergent, if present, is a nonionic detergent. Typical surfactants useful in the methods of the present invention include, but are not limited to, ammonium lauryl sulfate, polyethylene glycols, butyl glucoside, decyl glucoside, Polysorbate 80, lauric acid, myristic acid, palmitic acid, potassium palmitate, undecanoic acid, lauryl betaine, and lauryl alcohol. More preferably, the detergent, if present, is Triton X-100 and present in an amount of about 0.1% (w/v).

Non-specific binding of a labeled target nucleic acid to test compounds can be further minimized by treating the binding reaction with one or more blocking agents. In one embodiment, the binding reactions are treated with a blocking agent, *e.g.*, bovine serum albumin ("BSA"), before contacting with the labeled target nucleic acid. In another embodiment, the binding reactions are treated sequentially with at least two different blocking agents. This blocking step is preferably performed at room temperature for from about 0.5 to about 3 hours. In a subsequent step, the reaction mixture is further treated with unlabeled RNA having a mutation at the binding site. This blocking step is preferably performed at about 4°C for from about 12 hours to about 36 hours before addition of the dye-labeled target RNA. Preferably, the solution used in the one or more blocking steps is

substantially similar to the aqueous solution used to screen the library with the dye-labeled target RNA, *e.g.*, in pH and salt concentration.

Once contacted, the mixture of labeled target nucleic acid and the test compound is preferably maintained at 4°C for from about 1 day to about 5 days, preferably from about 2 days to about 3 days with constant agitation. To identify the reactions in which binding to the labeled target nucleic acid occurred, after the incubation period, bound from free compounds are determined using an electrophoretic technique (see Section 5.5.1), or any of the methods disclosed in Section 5.5 *infra*. In another embodiment, the complexed target nucleic acid does not need to be separated from the free target nucleic acid if a technique (*i.e.*, spectrometry) that differentiates between bound and unbound target nucleic acids is used.

The methods for identifying small molecules bound to labeled nucleic acid will vary with the type of label on the target nucleic acid. For example, if a target RNA is labeled with a visible of fluorescent dye, the target RNA complexes are preferably identified using a chromatographic technique that separates bound from free target by an electrophoretic or size differential technique using individual reactions. The reactions corresponding to changes in the migration of the complexed RNA can be cross-referenced to the small molecule compound(s) added to said reaction. Alternatively, complexed target RNA can be screened *en masse* and then separated from free target RNA using an electrophoretic or size differential technique, the resultant complexed target is then analyzed using a mass spectrometric technique. In this fashion the bound small molecule can be identified on the basis of its molecular weight. In this reaction *a priori* knowledge of the exact molecular weights of all compounds within the library is known. In another embodiment, the test compounds bound to the target nucleic acid may not require separation from the unbound target nucleic acid if a technique such as, but not limited to, spectrometry is used.

5.5. Separation Methods for Screening Test Compounds

Any method that detects an altered physical property of a target nucleic acid complexed to a test compound from the unbound target nucleic acid may be used for separation of the complexed and non-complexed target nucleic acids. Methods that can be utilized for the physical separation of complexed target RNA from unbound target RNA include, but are not limited to, electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography,

and nanoparticle aggregation.

5.5.1. Electrophoresis

5 Methods for separation of the complex of a target RNA bound to a test compound from the unbound RNA comprises any method of electrophoretic separation, including but not limited to, denaturing and non-denaturing polyacrylamide gel electrophoresis, urea gel electrophoresis, gel filtration, pulsed field gel electrophoresis, two dimensional gel electrophoresis, continuous flow electrophoresis, zone electrophoresis, agarose gel electrophoresis, and capillary electrophoresis.

10 In a preferred embodiment, an automated electrophoretic system comprising a capillary cartridge having a plurality of capillary tubes is used for high-throughput screening of test compounds bound to target RNA. Such an apparatus for performing automated capillary gel electrophoresis is disclosed in U.S. Patent Nos. 5,885,430; 15 5,916,428; 6,027,627; and 6,063,251, the disclosures of which are incorporated by reference in their entireties.

20 The device disclosed in U.S. Patent No. 5,885,430, which is incorporated by reference in its entirety, allows one to simultaneously introduce samples into a plurality of capillary tubes directly from microtiter trays having a standard size. U.S. Patent No. 5,885,430 discloses a disposable capillary cartridge which can be cleaned between 25 electrophoresis runs, the cartridge having a plurality of capillary tubes. A first end of each capillary tube is retained in a mounting plate, the first ends collectively forming an array in the mounting plate. The spacing between the first ends corresponds to the spacing between the centers of the wells of a microtiter tray having a standard size. Thus, the first ends of the capillary tubes can simultaneously be dipped into the samples present in the tray's wells. The cartridge is provided with a second mounting plate in which the second ends of the 30 capillary tubes are retained. The second ends of the capillary tubes are arranged in an array which corresponds to the wells in the microtiter tray, which allows for each capillary tube to be isolated from its neighbors and therefore free from cross-contamination, as each end is dipped into an individual well.

35 Plate holes may be provided in each mounting plate and the capillary tubes inserted through these plate holes. In such a case, the plate holes are sealed airtight so that the side of the mounting plate having the exposed capillary ends can be pressurized. Application of a positive pressure in the vicinity of the capillary openings in this mounting plate allows for the introduction of air and fluids during electrophoretic operations and also can be used to force out gel and other materials from the capillary tubes during

reconditioning. The capillary tubes may be protected from damage using a needle comprising a cannula and/or plastic tubes, and the like when they are placed in these plate holes. When metallic cannula or the like are used, they can serve as electrical contacts for current flow during electrophoresis. In the presence of a second mounting plate, the second mounting plate is provided with plate holes through which the second ends of the capillary tubes project. In this instance, the second mounting plate serves as a pressure containment member of a pressure cell and the second ends of the capillary tubes communicate with an internal cavity of the pressure cell. The pressure cell is also formed with an inlet and an outlet. Gels, buffer solutions, cleaning agents, and the like may be introduced into the internal cavity through the inlet, and each of these can simultaneously enter the second ends of the capillaries.

In another preferred embodiment, the automated electrophoretic system can comprise a chip system consisting of complex designs of interconnected channels that perform and analyze enzyme reactions using part of a channel design as a tiny, continuously operating electrophoresis material, where reactions with one sample are going on in one area of the chip while electrophoretic separation of the products of another sample is taking place in a different part of the chip. Such a system is disclosed in U.S. Patent Nos. 5,699,157; 5,842,787; 5,869,004; 5,876,675; 5,942,443; 5,948,227; 6,042,709; 6,042,710; 15 6,046,056; 6,048,498; 6,086,740; 6,132,685; 6,150,119; 6,150,180; 6,153,073; 6,167,910; 20 6,171,850; and 6,186,660, the disclosures of which are incorporated by reference in their entireties.

The system disclosed in U.S. Patent No. 5,699,157, which is hereby incorporated by reference in its entirety, provides for a microfluidic system for high-speed electrophoretic analysis of subject materials for applications in the fields of chemistry, 25 biochemistry, biotechnology, molecular biology and numerous other areas. The system has a channel in a substrate, a light source and a photoreceptor. The channel holds subject materials in solution in an electric field so that the materials move through the channel and separate into bands according to species. The light source excites fluorescent light in the species bands and the photoreceptor is arranged to receive the fluorescent light from the bands. The system further has a means for masking the channel so that the photoreceptor can receive the fluorescent light only at periodically spaced regions along the channel. The 30 system also has a unit connected to analyze the modulation frequencies of light intensity received by the photoreceptor so that velocities of the bands along the channel are determined, which allows the materials to be analyzed.

35 The system disclosed in U.S. Patent No. 5,699,157 also provides for a

method of performing high-speed electrophoretic analysis of subject materials, which comprises the steps of holding the subject materials in solution in a channel of a microfluidic system; subjecting the materials to an electric field so that the subject materials move through the channel and separate into species bands; directing light toward the channel; receiving light from periodically spaced regions along the channel simultaneously; and analyzing the frequencies of light intensity of the received light so that velocities of the bands along the channel can be determined for analysis of said materials. The determination of the velocity of a species band determines the electrophoretic mobility of the species and its identification.

U.S. Patent No. 5,842,787, which is hereby incorporated by reference in its entirety, is generally directed to devices and systems employ channels having, at least in part, depths that are varied over those which have been previously described (such as the device disclosed in U.S. Patent No. 5,699,157), wherein said channel depths provide numerous beneficial and unexpected results such as but not limited to, a reduction in sample perturbation, reduced non-specific sample mixture by diffusion, and increased resolution.

In another embodiment, the electrophoretic method of separation comprises polyacrylamide gel electrophoresis. In a preferred embodiment, the polyacrylamide gel electrophoresis is non-denaturing, so as to differentiate the mobilities of the target RNA bound to a test compound from free target RNA. If the polyacrylamide gel electrophoresis is denaturing, then the target RNA: test compound complex must be cross-linked prior to electrophoresis to prevent the disassociation of the target RNA from the test compound during electrophoresis. Such techniques are well known to one of skill in the art.

In one embodiment of the method, the binding of test compounds to target nucleic acid can be detected, preferably in an automated fashion, by gel electrophoretic analysis of interference footprinting. RNA can be degraded at specific base sites by enzymatic methods such as ribonucleases A, U₂, CL₃, T₁, Phy M, and *B. cereus* or chemical methods such as diethylpyrocarbonate, sodium hydroxide, hydrazine, piperidine formate, dimethyl sulfate, [2,12-dimethyl-3,7,11,17-tetraazacyclo[11.3.1]heptadeca-1(17),2,11,13,15-pentaenato] nickel(II) (NiCR), cobalt(II)chloride, or iron(II) ethylenediaminetetraacetate (Fe-EDTA) as described for example in Zheng *et al.*, 1999, Biochem. 37:2207-2214; Latham & Cech, 1989, Science 245:276-282; and Sambrook *et al.*, 2001, in Molecular Cloning: A Laboratory Manual, pp 12.61-12.73, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entireties. The

specific pattern of cleavage sites is determined by the accessibility of particular bases to the reagent employed to initiate cleavage and, as such, is therefore determined by the three-dimensional structure of the RNA.

5 The interaction of small molecules with a target nucleic acid can change the accessibility of bases to these cleavage reagents both by causing conformational changes in the target nucleic acid or by covering a base at the binding interface. When a test compound binds to the nucleic acid and changes the accessibility of bases to cleavage reagents, the observed cleavage pattern will change. This method can be used to identify and characterize the binding of small molecules to RNA as described, for example, by
10 Prudent *et al.*, 1995, J. Am. Chem. Soc. 117:10145-10146 and Mei *et al.*, 1998, Biochem. 37:14204-14212.

15 In the preferred embodiment of this technique, the detectably labeled target nucleic acid is incubated with an individual test compound and then subjected to treatment with a cleavage reagent, either enzymatic or chemical. The reaction mixture can be
20 preferably be examined directly, or treated further to isolate and concentrate the nucleic acid. The fragments produced are separated by electrophoresis and the pattern of cleavage can be compared to a cleavage reaction performed in the absence of test compound. A change in the cleavage pattern directly indicates that the test compound binds to the target nucleic acid. Multiple test compounds can be examined both in parallel and serially.

25 Other embodiments of electrophoretic separation include, but are not limited to urea gel electrophoresis, gel filtration, pulsed field gel electrophoresis, two dimensional gel electrophoresis, continuous flow electrophoresis, zone electrophoresis, and agarose gel electrophoresis.

25

5.5.2. Fluorescence Spectroscopy

30 In a preferred embodiment, fluorescence polarization spectroscopy, an optical detection method that can differentiate the proportion of a fluorescent molecule that is either bound or unbound in solution (e.g., the labeled target nucleic acid of the present invention), can be used to read reaction results without electrophoretic separation of the samples. Fluorescence polarization spectroscopy can be used to read the reaction results in the chip system disclosed in U.S. Patent Nos. 5,699,157; 5,842,787; 5,869,004; 5,876,675; 5,942,443; 5,948,227; 6,042,709; 6,042,710; 6,046,056; 6,048,498; 6,086,740; 6,132,685; 6,150,119; 6,150,180; 6,153,073; 6,167,910; 6,171,850; and 6,186,660, the disclosures of
35 which are incorporated by reference in their entireties. The application of fluorescence

polarization spectroscopy to the chip system disclosed in the U.S. Patents listed *supra* is fast, efficient, and well-adapted for high-throughput screening.

5 In another embodiment, a compound that has an affinity for the target nucleic acid of interest can be labeled with a fluorophore to screen for test compounds that bind to the target nucleic acid. For example, a pyrene-containing aminoglycoside analog was used to accurately monitor antagonist binding to a prokaryotic 16S rRNA A site (which comprises the natural target for aminoglycoside antibiotics) in a screen using a fluorescence quenching technique in a 96-well plate format (Hamasaki & Rando, 1998, *Anal. Biochem.* 10 261(2):183-90).

10 In another embodiment, fluorescence resonance energy transfer (FRET) can be used to screen for test compounds that bind to the target nucleic acid. FRET, a characteristic change in fluorescence, occurs when two fluorophores with overlapping emission and excitation wavelength bands are held together in close proximity, such as by a 15 binding event. In the preferred embodiment, the fluorophore on the target nucleic acid and the fluorophore on the test compounds will have overlapping excitation and emission spectra such that one fluorophore (the donor) transfers its emission energy to excite the other fluorophore (the acceptor). The acceptor preferably emits light of a different wavelength upon relaxing to the ground state, or relaxes non-radiatively to quench 20 fluorescence. FRET is very sensitive to the distance between the two fluorophores, and allows measurement of molecular distances less than 10 nm. For example, U.S. Patent 6,337,183 to Arenas *et al.*, which is incorporated by reference in its entirety, describes a screen for compounds that bind RNA that uses FRET to measure the effect of test 25 compounds on the stability of a target RNA molecule where the target RNA is labeled with both fluorescent acceptor and donor molecules and the distance between the two fluorophores as determined by FRET provides a measure of the folded structure of the RNA. Matsumoto *et al.* (2000, *Bioorg. Med. Chem. Lett.* 10:1857-1861) describe a system where a peptide that binds to HIV-1 TAR RNA is labeled on one end with a fluorescein 30 fluorophore and a tetramethylrhodamine on the other end. The conformational change of the peptide upon binding to the RNA provided a FRET signal to screen for compounds that bound to the TAR RNA.

35 In the preferred embodiment, both the target nucleic acid and a compound that has an affinity for the target nucleic acid of interest are labeled with fluorophores with overlapping emission and excitation spectra (donor and acceptor), including but not limited to fluorescein and derivatives, rhodamine and derivatives, cyanine dyes and derivatives, bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, pyrene, nanoparticles, or

non-fluorescent quenching molecules. Binding of a labeled test compound to the target nucleic acid can be identified by the change in observable fluorescence as a result of FRET.

If the target nucleic acid is labeled with the donor fluorophore, then the test compounds is labeled with the acceptor fluorophore. Conversely, if the target nucleic acid is labeled with the acceptor fluorophore, then the test compounds is labeled with the donor fluorophore. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions. The fluorophore on the target nucleic acid must be in close proximity to the binding site of the test compounds, but should not be incorporated into a target nucleic acid at the specific binding site at which test compounds are likely to bind, since the presence of a covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site.

In yet another embodiment, homogeneous time-resolved fluorescence (“HTRF”) techniques based on time-resolved energy transfer from lanthanide ion complexes to a suitable acceptor species can be adapted for high-throughput screening for inhibitors of RNA-protein complexes (Hemmilä, 1999, *J. Biomol. Screening* 4:303-307; Mathis, 1999, *J. Biomol. Screening* 4:309-313). HTRF is similar to fluorescence resonance energy transfer using conventional organic dye pairs, but has several advantages, such as increased sensitivity and efficiency, and background elimination (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356).

Fluorescence spectroscopy has traditionally been used to characterize DNA-protein and protein-protein interactions, but fluorescence spectroscopy has not been widely used to characterize RNA-protein interactions because of an interfering absorption of RNA nucleotides with the intrinsic tryptophan fluorescence of proteins (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356.). However, fluorescence spectroscopy has been used in studying the single tryptophan residue within the arginine-rich RNA-binding domain of Rev protein and its interaction with the RRE in a time-resolved fluorescence study (Kwon & Carson, 1998, *Anal. Biochem.* 264:133-140). Thus, in this invention, fluorescence spectroscopy is less preferred if the test compounds or peptides or proteins possess intrinsic tryptophan fluorescence. However, fluorescence spectroscopy can be used for test compounds that do not possess intrinsic fluorescence.

5.5.3. Surface Plasmon Resonance (“SPR”)

Surface plasmon resonance (SPR) can be used for determining kinetic rate constants and equilibrium constants for macromolecular interactions by following the

association project in "real time" (Schuck, 1997, *Annu. Rev. Biophys. Biomol. Struct.* 26:541-566).

The principle of SPR is summarized by Xavier *et al.* (*Trends Biotechnol.*, 2000, 18(8):349-356) as follows. Total internal reflection occurs at the boundary between 5 two substances of different refractive index. The incident light's electromagnetic field penetrates beyond the interface as an evanescent wave, which extends a few hundred nanometers beyond the surface into the medium. Insertion of a thin gold foil at the interface produced SPR owing to the absorption of the energy from the evanescent wave by free 10 electron clouds of the metal (plasmons). As a result of this absorbance, there is a drop in the intensity of the reflected light at a particular angle of incidence. The evanescent wave profile depends exquisitely on the refractive index of the medium it probes. Thus, the angle at which absorption occurs is very sensitive to the refractive changes in the external 15 medium. All proteins and nucleic acids are known to change the refractive index of water by a similar amount per unit mass, irrespective of their amino acid or nucleotide composition (the refractive index change is different for proteins and nucleic acids). When 20 the protein or nucleic acid content of the layer at the sensor changes, the refractive index also changes. Typically, one member of a complex is immobilized in a dextran layer and then the other member is introduced into the solution, either in a flow cell (Biacore AB, Uppsala, Sweden) or a stirred cuvette (Affinity Sensors, Santa Fe, New Mexico). It has 25 been determined that there is a linear correlation between the surface concentration of protein or nucleic acid and the shift in resonance angle, which can be used to quantitate kinetic rate constants and/or the equilibrium constants.

In the present invention, the target RNA may be immobilized to the sensor 25 surface through a streptavidin-biotin linkage, the linkage of which is disclosed by Crouch *et al.* (*Methods Mol. Biol.*, 1999, 118:143-160). The RNA is biotinylated either during synthesis or post-synthetically via the conversion of the 3' terminal ribonucleoside of the RNA into a reactive free amino group or using a T7 polymerase incorporated guanosine 30 monophosphorothioate at the 5' end. SPR has been used to determine the stoichiometry and affinity of the interaction between the HIV Rev protein and the RRE (Van Ryk & Venkatesan, 1999, *J. Biol. Chem.* 274:17452-17463) and the aminoglycoside antibiotics 35 with RRE and a model RNA derived from the 16S ribosomal A site, respectively (Hendrix *et al.*, 1997, *J. Am. Chem. Soc.* 119:3641-3648; Wong *et al.*, 1998, *Chem. Biol.* 5:397-406).

In one embodiment of the present invention, the target nucleic acid can be 35 immobilized to a sensor surface (e.g., by a streptavidin-biotin linkage) and SPR can be used

to (a) determine whether the target RNA binds a test compound and (b) further characterize the binding of the target nucleic acids of the present invention to a test compound.

5 5.5.4. Mass Spectrometry

An automated method for analyzing mass spectrometer data which can analyze complex mixtures containing many thousands of components and can correct for background noise, multiply charged peaks and atomic isotope peaks is described in U.S. Patent No. 6,147,344, which is hereby incorporated by reference in its entirety. The system disclosed in U.S. Patent No. 6,147,344 is a method for analyzing mass spectrometer data in which a control sample measurement is performed providing a background noise check. The peak height and width values at each m/z ratio as a function of time are stored in a memory. A mass spectrometer operation on a material to be analyzed is performed and the peak height and width values at each m/z ratio versus time are stored in a second memory location. The mass spectrometer operation on the material to be analyzed is repeated a fixed number of times and the stored control sample values at each m/z ratio level at each time increment are subtracted from each corresponding one from the operational runs, thus producing a difference value at each mass ratio for each of the multiple runs at each time increment. If the MS value minus the background noise does not exceed a preset value, the m/z ratio data point is not recorded, thus eliminating background noise, chemical noise and false positive peaks from the mass spectrometer data. The stored data for each of the multiple runs is then compared to a predetermined value at each m/z ratio and the resultant series of peaks, which are now determined to be above the background, is stored in the m/z points in which the peaks are of significance.

25 One possibility for the utilization of mass spectrometry in high throughput screening is the integration of SPR with mass spectrometry. Approaches that have been tried are direct analysis of the analyte retained on the sensor chip and mass spectrometry with the eluted analyte (Sonksen *et al.*, 1998, *Anal. Chem.* 70:2731-2736; Nelson & Krone, 1999, *J. Mol. Recog.* 12:77-93). Further developments, especially in the interfacing of the sensor chip with the mass spectrometer and in reusing the sensor chip, are required to make 30 SPR combined with mass spectroscopy a high-throughput method for biomolecular interaction analysis and the screening of targets for small molecule inhibitors (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356).

35 In one embodiment of the present invention, the target nucleic acid complexed to a test compound can be determined by any of the mass spectrometry processed described *supra*. Furthermore, mass spectrometry can also be used to elucidate

the structure of the test compound.

5.5.5. Scintillation Proximity Assay (“SPA”)

Scintillation Proximity Assay (“SPA”) is a method that can be used for screening small molecules that bind to the target RNAs. SPA would involve radiolabeling either the target RNA or the test compound and then quantitating its binding to the other member to a bead or a surface impregnated with a scintillant (Cook, 1996, *Drug Discov. Today* 1:287-294). Currently, fluorescence-based techniques are preferred for high-throughput screening (Pope *et al.*, 1999, *Drug Discov. Today* 4:350-362).

Screening for small molecules that inhibit Tat peptide:TAR RNA interaction has been performed with SPA, and inhibitors of the interaction were isolated and characterized (Mei *et al.*, 1997, *Bioorg. Med. Chem.* 5:1173-1184; Mei *et al.*, 1998, *Biochemistry* 37:14204-14212). A similar approach can be used to identify small molecules that directly bind to a preselected target RNA element in accordance with the invention can be utilized.

SPA can be adapted to high throughput screening by the availability of microplates, wherein the scintillant is directly incorporated into the plastic of the microtiter wells (Nakayama *et al.*, 1998, *J. Biomol. Screening* 3:43-48). Thus, one embodiment of the present invention comprises (a) labeling of the target nucleic acid with a radioactive or fluorescent label; (b) contacted the labeled nucleic acid with test compounds, wherein each test compound is in a microtiter well coated with scintillant and is tethered to the microtiter well; and (c) identifying and quantifying the test compounds bound to the target nucleic acid with SPA, wherein the test compound is identified by virtue of its location in the microplate.

5.5.6. Structure-Activity Relationships (“SAR”) by NMR Spectroscopy

NMR spectroscopy is a valuable technique for identifying complexed target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects, and NMR-based approaches have been used in the identification of small molecule binders of protein drug targets (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356). The determination of structure-activity relationships (“SAR”) by NMR is the first method for NMR described in which small molecules that bind adjacent subsites are identified by two-dimensional ^1H - ^{15}N spectra of the target protein (Shuker *et al.*, 1996, *Science* 274:1531-1534). The signal from the bound molecule is monitored by employing line broadening, transferred NOEs and pulsed field gradient

diffusion measurements (Moore, 1999, *Curr. Opin. Biotechnol.* 10:54-58). A strategy for lead generation by NMR using a library of small molecules has been recently described (Fejzo *et al.*, 1999, *Chem. Biol.* 6:755-769).

5 In one embodiment of the present invention, the target nucleic acid complexed to a test compound can be determined by SAR by NMR. Furthermore, SAR by NMR can also be used to elucidate the structure of the test compound.

5.5.7. Size Exclusion Chromatography

10 In another embodiment of the present invention, size-exclusion chromatography is used to purify test compounds that are bound to a target nucleic acid from a complex mixture of compounds. Size-exclusion chromatography separates molecules based on their size and uses gel-based media comprised of beads with specific size distributions. When applied to a column, this media settles into a tightly packed matrix and forms a complex array of pores. Separation is accomplished by the inclusion or exclusion 15 of molecules by these pores based on molecular size. Small molecules are included into the pores and, consequently, their migration through the matrix is retarded due to the added distance they must travel before elution. Large molecules are excluded from the pores and migrate with the void volume when applied to the matrix. In the present invention, a target nucleic acid is incubated with a mixture of test compounds while free in solution and 20 allowed to reach equilibrium. When applied to a size exclusion column, test compounds free in solution are retained by the column, and test compounds bound to the target nucleic acid are passed through the column. In a preferred embodiment, spin columns commonly used for "desalting" of nucleic acids will be employed to separate bound from unbound test 25 compounds (e.g., Bio-Spin columns manufactured by BIO-RAD). In another embodiment, the size exclusion matrix is packed into multiwell plates to allow high throughput separation of mixtures (e.g., PLASMID 96-well SEC plates manufactured by Millipore).

5.5.8. Affinity Chromatography

30 In one embodiment of the present invention, affinity capture is used to purify test compounds that are bound to a target nucleic acid labeled with an affinity tag from a complex mixture of compounds. To accomplish this, a target nucleic acid labeled with an affinity tag is incubated with a mixture of test compounds while free in solution and then captured to a solid support once equilibrium has been established; alternatively, target 35 nucleic acids labeled with an affinity tag can be captured to a solid support first and then allowed to reach equilibrium with a mixture of test compounds.

5 The solid support is typically comprised of, but not limited to, cross-linked agarose beads that are coupled with a ligand for the affinity tag. Alternatively, the solid support may be a glass, silicon, metal, or carbon, plastic (polystyrene, polypropylene) surface with or without a self-assembled monolayer (SAM) either with a covalently attached ligand for the affinity tag, or with inherent affinity for the tag on the target nucleic acid.

10 Once the complex between the target nucleic acid and test compound has reached equilibrium and has been captured, one skilled in the art will appreciate that the retention of bound compounds and removal of unbound compounds is facilitated by washing the solid support with large excesses of binding reaction buffer. Furthermore, retention of high affinity compounds and removal of low affinity compounds can be accomplished by a number of means that increase the stringency of washing; these means include, but are not limited to, increasing the number and duration of washes, raising the salt concentration of the wash buffer, addition of detergent or surfactant to the wash buffer, and addition of non-specific competitor to the wash buffer.

15

20 In one embodiment, the test compounds themselves are detectably labeled with fluorescent dyes, radioactive isotopes, or nanoparticles. When the test compounds are applied to the captured target nucleic acid in a spatially addressed fashion (e.g., in separate wells of a 96-well microplate), binding between the test compounds and the target nucleic acid can be determined by the presence of the detectable label on the test compound using fluorescence.

25 Following the removal of unbound compounds, bound compounds with high affinity for the target nucleic acid can be eluted from the immobilized target nucleic acids and analyzed. The elution of test compounds can be accomplished by any means that break the non-covalent interactions between the target nucleic acid and compound. Means for elution include, but are not limited to, changing the pH, changing the salt concentration, the application of organic solvents, and the application of molecules that compete with the bound ligand. In a preferred embodiment, the means employed for elution will release the compound from the target RNA, but will not effect the interaction between the affinity tag and the solid support, thereby achieving selective elution of test compound. Moreover, a preferred embodiment will employ an elution buffer that is volatile to allow for subsequent concentration by lyophilization of the eluted compound (e.g., 0 M to 5 M ammonium acetate).

30

5.5.9. Nanoparticle Aggregation

In one embodiment of the present invention, both the target nucleic acid and the test compounds are labeled with nanoparticles. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to Ag₂S, ZnS, CdS, CdTe, Au, or TiO₂. Methods for the attachment of nucleic acids and small molecules to nanoparticles are well known to one of skill in the art (reviewed in Niemeyer, 2001, *Angew. Chem. Int. Ed.* 40:4129-4158. The references cited therein are hereby incorporated by reference in their entireties). In particular, if multiple copies of the target nucleic acid are attached to a single nanoparticle and multiple copies of a test compound are attached to another nanoparticle, then interaction between the test compound and target nucleic acid will induce aggregation of nanoparticles as described, for example, by Mitchel *et al.* 1999, *J. Am. Chem. Soc.* 121:8122-8123. The aggregate can be detected by changes in absorbance or fluorescence spectra and physically separated from the unbound components through filtration or centrifugation.

5.6. Methods for Identifying or Characterizing the Test Compounds Bound to the Target Nucleic Acids

If the library comprises arrays or microarrays of test compounds, wherein each test compound has an address or identifier, the test compound can be deconvoluted, *e.g.*, by cross-referencing the positive sample to original compound list that was applied to the individual test assays.

If the library is a peptide or nucleic acid library, the sequence of the test compound can be determined by direct sequencing of the peptide or nucleic acid. Such methods are well known to one of skill in the art.

A number of physico-chemical techniques can be used for the *de novo* characterization of test compounds bound to the target.

5.6.1. Mass Spectrometry

Mass spectrometry (*e.g.*, electrospray ionization ("ESI") and matrix-assisted laser desorption-ionization ("MALDI"), Fourier-transform ion cyclotron resonance ("FT-ICR")) can be used both for high-throughput screening of test compounds that bind to a target RNA and elucidating the structure of the test compound. Thus, one example of mass spectroscopy is that separation of a bound and unbound complex and test compound structure elucidation can be carried out in a single step.

5 MALDI uses a pulsed laser for desorption of the ions and a time-of-flight analyzer, and has been used for the detection of noncovalent tRNA:amino-acyl-tRNA synthetase complexes (Gruic-Sovulj *et al.*, 1997, *J. Biol. Chem.* 272:32084-32091). However, covalent cross-linking between the target nucleic acid and the test compound is required for detection, since a non-covalently bound complex may dissociate during the MALDI process.

10 ESI mass spectrometry ("ESI-MS") has been of greater utility for studying non-covalent molecular interactions because, unlike the MALDI process, ESI-MS generates molecular ions with little to no fragmentation (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356). ESI-MS has been used to study the complexes formed by HIV Tat peptide and protein with the TAR RNA (Sannes-Lowery *et al.*, 1997, *Anal. Chem.* 69:5130-5135).

15 Fourier-transform ion cyclotron resonance ("FT-ICR") mass spectrometry provides high-resolution spectra, isotope-resolved precursor ion selection, and accurate mass assignments (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356). FT-ICR has been used to study the interaction of aminoglycoside antibiotics with cognate and non-cognate RNAs (Hofstadler *et al.*, 1999, *Anal. Chem.* 71:3436-3440; Griffey *et al.*, 1999, *Proc. Natl. Acad. Sci. USA* 96:10129-10133). As true for all of the mass spectrometry methods discussed herein, FT-ICR does not require labeling of the target RNA or a test compound.

20 An advantage of mass spectroscopy is not only the elucidation of the structure of the test compound, but also the determination of the structure of the test compound bound to the preselected target RNA. Such information can enable the discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

25 **5.6.2. NMR Spectroscopy**
As described above, NMR spectroscopy is a technique for identifying binding sites in target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects. Examples of NMR that can be used for the invention include, but are not limited to, one-dimentional NMR, two-dimentional NMR, correlation spectroscopy ("COSY"), and nuclear Overhauser effect ("NOE") spectroscopy. Such methods of structure determination of test compounds are well known to one of skill in the art.

30 35 Similar to mass spectroscopy, an advantage of NMR is the not only the elucidation of the structure of the test compound, but also the determination of the structure

of the test compound bound to the preselected target RNA. Such information can enable the discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

5

5.6.3. Vibrational Spectroscopy

Vibrational spectroscopy (e.g. infrared (IR) spectroscopy or Raman spectroscopy) can be used for elucidating the structure of the test compound on the isolated bead.

10

Infrared spectroscopy measures the frequencies of infrared light (wavelengths from 100 to 10,000 nm) absorbed by the test compound as a result of excitation of vibrational modes according to quantum mechanical selection rules which require that absorption of light cause a change in the electric dipole moment of the molecule. The infrared spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

15

Infrared spectra can be measured in a scanning mode by measuring the absorption of individual frequencies of light, produced by a grating which separates frequencies from a mixed-frequency infrared light source, by the test compound relative to a standard intensity (double-beam instrument) or pre-measured ('blank') intensity (single-beam instrument). In a preferred embodiment, infrared spectra are measured in a pulsed mode (FT-IR) where a mixed beam, produced by an interferometer, of all infrared light frequencies is passed through or reflected off the test compound. The resulting interferogram, which may or may not be added with the resulting interferograms from subsequent pulses to increase the signal strength while averaging random noise in the electronic signal, is mathematically transformed into a spectrum using Fourier Transform or Fast Fourier Transform algorithms.

20

Raman spectroscopy measures the difference in frequency due to absorption of infrared frequencies of scattered visible or ultraviolet light relative to the incident beam. The incident monochromatic light beam, usually a single laser frequency, is not truly absorbed by the test compound but interacts with the electric field transiently. Most of the light scattered off the sample will be unchanged (Rayleigh scattering) but a portion of the scatter light will have frequencies that are the sum or difference of the incident and molecular vibrational frequencies. The selection rules for Raman (inelastic) scattering require a change in polarizability of the molecule. While some vibrational transitions are observable in both infrared and Raman spectrometry, most are observable only with one or

25

30

35

the other technique. The Raman spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

5 Raman spectra are measured by submitting monochromatic light to the sample, either passed through or preferably reflected off, filtering the Rayleigh scattered light, and detecting the frequency of the Raman scattered light. An improved Raman spectrometer is described in US Patent No. 5,786,893 to Fink *et al.*, which is hereby incorporated by reference.

10 Vibrational microscopy can be measured in a spatially resolved fashion to address single beads by integration of a visible microscope and spectrometer. A microscopic infrared spectrometer is described in U.S. Patent No. 5,581,085 to Reffner *et al.*, which is hereby incorporated by reference in its entirety. An instrument that simultaneously performs a microscopic infrared and microscopic Raman analysis on a 15 sample is described in U.S. Patent No. 5,841,139 to Sostek *et al.*, which is hereby incorporated by reference in its entirety.

20 In the preferred embodiment, test compounds can be identified by matching the IR or Raman spectra of a test compound to a dataset of vibrational (IR or Raman) spectra previously acquired for each compound in the combinatorial library. By this method, the spectra of compounds with known structure are recorded so that comparison 25 with these spectra can identify compounds again when isolated from RNA binding experiments.

5.7. Secondary Biological Screens

25 The test compounds identified in the binding assay (for convenience referred to herein as a "lead" compound) can be tested for biological activity using host cells containing or engineered to contain the target RNA element coupled to a functional readout system. For example, the lead compound can be tested in a host cell engineered to contain the target RNA element controlling the expression of a reporter gene. In this example, the 30 lead compounds are assayed in the presence or absence of the target RNA. Alternatively, a phenotypic or physiological readout can be used to assess activity of the target RNA in the presence and absence of the lead compound.

35 In one embodiment, the lead compound can be tested in a host cell engineered to contain the target RNA element controlling the expression of a reporter gene, such as, but not limited to, β -galactosidase, green fluorescent protein, red fluorescent protein, luciferase, chloramphenicol acetyltransferase, alkaline phosphatase, and β -

lactamase. In a preferred embodiment, a cDNA encoding the target element is fused upstream to a reporter gene wherein translation of the reporter gene is repressed upon binding of the lead compound to the target RNA. In other words, the steric hindrance caused by the binding of the lead compound to the target RNA repressed the translation of the reporter gene. This method, termed the translational repression assay procedure ("TRAP") has been demonstrated in *E. coli* and *S. cerevisiae* (Jain & Belasco, 1996, *Cell* 87(1):115-25; Huang & Schreiber, 1997, *Proc. Natl. Acad. Sci. USA* 94:13396-13401).

5 In another embodiment, a phenotypic or physiological readout can be used to assess activity of the target RNA in the presence and absence of the lead compound. For 10 example, the target RNA may be overexpressed in a cell in which the target RNA is endogenously expressed. Where the target RNA controls expression of a gene product involved in cell growth or viability, the *in vivo* effect of the lead compound can be assayed by measuring the cell growth or viability of the target cell. Alternatively, a reporter gene 15 can also be fused downstream of the target RNA sequence and the effect of the lead compound on reporter gene expression can be assayed.

15 Alternatively, the lead compounds identified in the binding assay can be tested for biological activity using animal models for a disease, condition, or syndrome of interest. These include animals engineered to contain the target RNA element coupled to a 20 functional readout system, such as a transgenic mouse. Animal model systems can also be used to demonstrate safety and efficacy.

25 Compounds displaying the desired biological activity can be considered to be lead compounds, and will be used in the design of congeners or analogs possessing useful pharmacological activity and physiological profiles. Following the identification of a lead compound, molecular modeling techniques can be employed, which have proven to be useful in conjunction with synthetic efforts, to design variants of the lead that can be more effective. These applications may include, but are not limited to, Pharmacophore Modeling (cf. Lamothe, *et al.* 1997, *J. Med. Chem.* 40: 3542; Mottola *et al.* 1996, *J. Med. Chem.* 39: 285; Beusen *et al.* 1995, *Biopolymers* 36: 181; P. Fossa *et al.* 1998, *Comput. Aided Mol. Des.* 12: 361), QSAR development (cf. Siddiqui *et al.* 1999, *J. Med. Chem.* 42: 4122; 30 Barreca *et al.* 1999 *Bioorg. Med. Chem.* 7: 2283; Kroemer *et al.* 1995, *J. Med. Chem.* 38: 4917; Schaal *et al.* 2001, *J. Med. Chem.* 44: 155; Buolamwini & Assefa 2002, *J. Mol. Chem.* 45: 84), Virtual docking and screening/scoring (cf. Anzini *et al.* 2001, *J. Med. Chem.* 44: 1134; Faaland *et al.* 2000, *Biochem. Cell. Biol.* 78: 415; Silvestri *et al.* 2000, *Bioorg. Med. Chem.* 8: 2305; J. Lee *et al.* 2001, *Bioorg. Med. Chem.* 9: 19), and Structure 35 Prediction using RNA structural programs including, but not limited to mFold (as described

5 by Zuker *et al.* Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology pp. 11-43, J. Barciszewski & B.F.C. Clark, eds. (NATO ASI Series, Kluwer Academic Publishers, 1999) and Mathews *et al.* 1999 J. Mol. Biol. 288: 911-940); RNAmotif (Macke *et al.* 2001, Nucleic Acids Res. 29: 4724-4735; and the Vienna RNA package (Hofacker *et al.* 1994, Monatsh. Chem. 125: 167-188).

10 Further examples of the application of such techniques can be found in several review articles, such as Rotivinen *et al.*, 1988, Acta Pharmaceutical Fennica 97:159-166; Ripka, 1998, New Scientist 54-57; McKinlay & Rossmann, 1989, Annu. Rev. Pharmacol. Toxicol. 29:111-122; Perry & Davies, QSAR: Quantitative Structure-Activity Relationships in Drug Design pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis & Dean, 1989, Proc. R. Soc. Lond. 236:125-140 and 141-162; Askew *et al.*, 1989, J. Am. Chem. Soc. 111:1082-1090. Molecular modeling tools employed may include those from Tripos, Inc., 15 St. Louis, Missouri (e.g., Sybyl/UNITY, CONCORD, DiverseSolutions), Accelrys, San Diego, California (e.g., Catalyst, Wisconsin Package {BLAST, etc.}), Schrodinger, Portland, Oregon (e.g., QikProp, QikFit, Jaguar) or other such vendors as BioDesign, Inc. (Pasadena, California), Allelix, Inc. (Mississauga, Ontario, Canada), and Hypercube, Inc. (Cambridge, Ontario, Canada), and may include privately designed and/or "academic" 20 software (e.g. RNAMotif, mFOLD). These application suites and programs include tools for the atomistic construction and analysis of structural models for drug-like molecules, proteins, and DNA or RNA and their potential interactions. They also provide for the calculation of important physical properties, such as solubility estimates, permeability metrics, and empirical measures of molecular "druggability" (e.g., Lipinski "Rule of 5" as described by Lipinski *et al.* 1997, Adv. Drug Delivery Rev. 23: 3-25). Most importantly, 25 they provide appropriate metrics and statistical modeling power (such as the patented CoMFA technology in Sybyl as described in US Patents 6,240,374 and 6,185,506) to develop Quantitative Structural Activity Relationships (QSARs) which are used to guide the synthesis of more efficacious clinical development candidates while improving 30 desirable physical properties, as determined by results from the aforementioned secondary screening protocols.

5.8. Use of Identified Compounds That Bind RNA to Treat/Prevent Disease

35 Biologically active compounds identified using the methods of the invention or a pharmaceutically acceptable salt thereof can be administered to a patient, preferably a mammal, more preferably a human, suffering from a disease whose progression is

associated with a target RNA:host cell factor interaction *in vivo*. In certain embodiments, such compounds or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with an RNA:host cell factor interaction *in vivo*.

5 In one embodiment, "treatment" or "treating" refers to an amelioration of a disease, or at least one discernible symptom thereof. In another embodiment, "treatment" or "treating" refers to an amelioration of at least one measurable physical parameter, not necessarily discernible by the patient. In yet another embodiment, "treatment" or "treating" refers to inhibiting the progression of a disease, either physically, *e.g.*, stabilization of a 10 discernible symptom, physiologically, *e.g.*, stabilization of a physical parameter, or both. In yet another embodiment, "treatment" or "treating" refers to delaying the onset of a disease.

15 In certain embodiments, the compound or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with an RNA:host cell factor interaction *in vivo*. As used herein, "prevention" or "preventing" refers to a reduction of the risk of acquiring a disease. In one embodiment, the compound or a pharmaceutically acceptable salt thereof is administered as a preventative measure to a patient. According to this 20 embodiment, the patient can have a genetic predisposition to a disease, such as a family history of the disease, or a non-genetic predisposition to the disease. Accordingly, the compound and pharmaceutically acceptable salts thereof can be used for the treatment of one manifestation of a disease and prevention of another.

25 When administered to a patient, the compound or a pharmaceutically acceptable salt thereof is preferably administered as component of a composition that optionally comprises a pharmaceutically acceptable vehicle. The composition can be administered orally, or by any other convenient route, for example, by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (*e.g.*, oral mucosa, rectal, and intestinal mucosa, *etc.*) and may be administered together with another 30 biologically active agent. Administration can be systemic or local. Various delivery systems are known, *e.g.*, encapsulation in liposomes, microparticles, microcapsules, capsules, *etc.*, and can be used to administer the compound and pharmaceutically acceptable salts thereof.

35 Methods of administration include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, oral, sublingual, intranasal, intracerebral, intravaginal, transdermal, rectally, by inhalation, or topically, particularly to the ears, nose, eyes, or skin. The mode of administration is left to

the discretion of the practitioner. In most instances, administration will result in the release of the compound or a pharmaceutically acceptable salt thereof into the bloodstream.

5 In specific embodiments, it may be desirable to administer the compound or a pharmaceutically acceptable salt thereof locally. This may be achieved, for example, and not by way of limitation, by local infusion during surgery, topical application, e.g., in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers.

10 15 In certain embodiments, it may be desirable to introduce the compound or a pharmaceutically acceptable salt thereof into the central nervous system by any suitable route, including intraventricular, intrathecal and epidural injection. Intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir.

15 20 Pulmonary administration can also be employed, e.g., by use of an inhaler or nebulizer, and formulation with an aerosolizing agent, or via perfusion in a fluorocarbon or synthetic pulmonary surfactant. In certain embodiments, the compound and pharmaceutically acceptable salts thereof can be formulated as a suppository, with traditional binders and vehicles such as triglycerides.

20 25 In another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a vesicle, in particular a liposome (see Langer, 1990, *Science* 249:1527-1533; Treat *et al.*, in *Liposomes in the Therapy of Infectious Disease and Cancer*, Lopez-Berestein and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, *ibid.*, pp. 317-327; see generally *ibid.*).

25 30 35 In yet another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a controlled release system (see, e.g., Goodson, in *Medical Applications of Controlled Release*, *supra*, vol. 2, pp. 115-138 (1984)). Other controlled-release systems discussed in the review by Langer, 1990, *Science* 249:1527-1533) may be used. In one embodiment, a pump may be used (see Langer, *supra*; Sefton, 1987, *CRC Crit. Ref. Biomed. Eng.* 14:201; Buchwald *et al.*, 1980, *Surgery* 88:507; Saudek *et al.*, 1989, *N. Engl. J. Med.* 321:574). In another embodiment, polymeric materials can be used (see *Medical Applications of Controlled Release*, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); *Controlled Drug Bioavailability, Drug Product Design and Performance*, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, 1983, *J. Macromol. Sci. Rev. Macromol. Chem.* 23:61; see also Levy *et al.*, 1985, *Science*

228:190; During *et al.*, 1989, Ann. Neurol. 25:351; Howard *et al.*, 1989, J. Neurosurg. 71:105). In yet another embodiment, a controlled-release system can be placed in proximity of a target RNA of the compound or a pharmaceutically acceptable salt thereof, thus requiring only a fraction of the systemic dose.

5 Compositions comprising the compound or a pharmaceutically acceptable salt thereof ("compound compositions") can additionally comprise a suitable amount of a pharmaceutically acceptable vehicle so as to provide the form for proper administration to the patient.

10 In a specific embodiment, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, mammals, and more particularly in humans. The term "vehicle" refers to a diluent, adjuvant, excipient, or carrier with which a compound of the invention is administered. Such pharmaceutical vehicles can be liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. The pharmaceutical vehicles can be saline, gum acacia, gelatin, starch paste, talc, keratin, colloidal silica, urea, and the like. In addition, auxiliary, stabilizing, thickening, lubricating and coloring agents may be used. When administered to a patient, the pharmaceutically acceptable vehicles are preferably sterile. Water is a preferred vehicle when the compound of the invention is administered intravenously. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid vehicles, particularly for injectable solutions. Suitable pharmaceutical vehicles also include excipients such as starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. Compound compositions, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents.

15 Compound compositions can take the form of solutions, suspensions, emulsion, tablets, pills, pellets, capsules, capsules containing liquids, powders, sustained-release formulations, suppositories, emulsions, aerosols, sprays, suspensions, or any other form suitable for use. In one embodiment, the pharmaceutically acceptable vehicle is a capsule (see e.g., U.S. Patent No. 5,698,155). Other examples of suitable pharmaceutical vehicles are described in Remington's Pharmaceutical Sciences, Alfonso R. Gennaro, ed., Mack Publishing Co. Easton, PA, 19th ed., 1995, pp. 1447 to 1676, incorporated herein by reference.

In a preferred embodiment, the compound or a pharmaceutically acceptable salt thereof is formulated in accordance with routine procedures as a pharmaceutical composition adapted for oral administration to human beings. Compositions for oral delivery may be in the form of tablets, lozenges, aqueous or oily suspensions, granules, 5 powders, emulsions, capsules, syrups, or elixirs, for example. Orally administered compositions may contain one or more agents, for example, sweetening agents such as fructose, aspartame or saccharin; flavoring agents such as peppermint, oil of wintergreen, or cherry; coloring agents; and preserving agents, to provide a pharmaceutically palatable 10 preparation. Moreover, where in tablet or pill form, the compositions can be coated to delay disintegration and absorption in the gastrointestinal tract thereby providing a sustained action over an extended period of time. Selectively permeable membranes 15 surrounding an osmotically active driving compound are also suitable for orally administered compositions. In these later platforms, fluid from the environment surrounding the capsule is imbibed by the driving compound, which swells to displace the agent or agent composition through an aperture. These delivery platforms can provide an essentially zero order delivery profile as opposed to the spiked profiles of immediate 20 release formulations. A time delay material such as glycerol monostearate or glycerol stearate may also be used. Oral compositions can include standard vehicles such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium 25 carbonate, and the like. Such vehicles are preferably of pharmaceutical grade. Typically, compositions for intravenous administration comprise sterile isotonic aqueous buffer. 30 Where necessary, the compositions may also include a solubilizing agent.

In another embodiment, the compound or a pharmaceutically acceptable salt thereof can be formulated for intravenous administration. Compositions for intravenous 25 administration may optionally include a local anesthetic such as lignocaine to lessen pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water-free 30 concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the compound or a pharmaceutically acceptable salt thereof is to be administered by infusion, it can be dispensed, for example, with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the compound or a 35 pharmaceutically acceptable salt thereof is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

The amount of a compound or a pharmaceutically acceptable salt thereof that will be effective in the treatment of a particular disease will depend on the nature of the disease, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The 5 precise dose to be employed will also depend on the route of administration, and the seriousness of the disease, and should be decided according to the judgment of the practitioner and each patient's circumstances. However, suitable dosage ranges for oral administration are generally about 0.001 milligram to about 200 milligrams of a compound or a pharmaceutically acceptable salt thereof per kilogram body weight per day. In specific 10 preferred embodiments of the invention, the oral dose is about 0.01 milligram to about 100 milligrams per kilogram body weight per day, more preferably about 0.1 milligram to about 75 milligrams per kilogram body weight per day, more preferably about 0.5 milligram to 5 milligrams per kilogram body weight per day. The dosage amounts described herein refer 15 to total amounts administered; that is, if more than one compound is administered, or if a compound is administered with a therapeutic agent, then the preferred dosages correspond to the total amount administered. Oral compositions preferably contain about 10% to about 95% active ingredient by weight.

Suitable dosage ranges for intravenous (i.v.) administration are about 0.01 20 milligram to about 100 milligrams per kilogram body weight per day, about 0.1 milligram to about 35 milligrams per kilogram body weight per day, and about 1 milligram to about 10 milligrams per kilogram body weight per day. Suitable dosage ranges for intranasal administration are generally about 0.01 pg/kg body weight per day to about 1 mg/kg body 25 weight per day. Suppositories generally contain about 0.01 milligram to about 50 milligrams of a compound of the invention per kilogram body weight per day and comprise active ingredient in the range of about 0.5% to about 10% by weight.

Recommended dosages for intradermal, intramuscular, intraperitoneal, 30 subcutaneous, epidural, sublingual, intracerebral, intravaginal, transdermal administration or administration by inhalation are in the range of about 0.001 milligram to about 200 milligrams per kilogram of body weight per day. Suitable doses for topical administration are in the range of about 0.001 milligram to about 1 milligram, depending on the area of administration. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems. Such animal models and systems are well 35 known in the art.

The compound and pharmaceutically acceptable salts thereof are preferably assayed *in vitro* and *in vivo*, for the desired therapeutic or prophylactic activity, prior to use

in humans. For example, *in vitro* assays can be used to determine whether it is preferable to administer the compound, a pharmaceutically acceptable salt thereof, and/or another therapeutic agent. Animal model systems can be used to demonstrate safety and efficacy.

5 A variety of compounds can be used for treating or preventing diseases in mammals. Types of compounds include, but are not limited to, peptides, peptide analogs including peptides comprising non-natural amino acids, *e.g.*, D-amino acids, phosphorous analogs of amino acids, such as α -amino phosphonic acids and α -amino phosphinic acids, or amino acids having non-peptide linkages, nucleic acids, nucleic acid analogs such as phosphorothioates or peptide nucleic acids ("PNAs"), hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose.

10

15 6. EXAMPLE: THERAPEUTIC TARGETS

The therapeutic targets presented herein are by way of example, and the present invention is not to be limited by the targets described herein. The therapeutic targets presented herein as DNA sequences are understood by one of skill in the art that the sequences can be converted to RNA sequences.

20

6.1. Tumor Necrosis Factor Alpha ("TNF- α ")

GenBank Accession # X01394:

1 gcagaggacc agctaagagg gagagaagca actacagacc cccctgaaa acaaccctca
61 gacgccacat cccctgacaa gctgccaggc aggttcttt ccttcacat actgaccac
121 ggctccaccc tctctccctt gaaaggaca ccatgagcac tgaaagcatg atccggacg
181 tggagctggc cgaggaggcg ctcccaaga agacaggggg gccccaggc tccaggcggt
241 gcttgttctt cagcctttc tccttcttga tggcaggcagg cgccaccacg ctcttctgccc
301 tgctgcactt tggagtgtatc ggcccccaga ggaaagatgg ccccaaggatc ctctcttcaa
361 tcagccctt ggcaggca gtcagatcat ttctcgaac cccgagtgac aagctgttag
421 cccatgttgtt agcaaaccctt caagctgagg ggcagatcca gtggctgaac cgccggccca
481 atgcccttctt ggccaatggc gtggagctga gagataacca gctgggttg ccacatcagg
541 gcctgtacctt catctactcc caggccctt tcaagggcca aggctcccccc tccaccatg
601 tgcttcac ccacaccatc agccgcattcg ccgttcttta ccagaccaag gtcaacctcc
661 tctctgccc caagagcccc tggcaggagg agaccccaaga gggggctgag gccaaggccct
721 ggtatgagcc catctatctg ggaggggtct tccagcttggaa gaagggtgac cgactcagcg
781 ctgagatcaa tcggccgac tatctcgact ttggcaggc tggcaggc tactttggga

841 tcattgccct gtgaggagga cgaacatcca accttcccaa acgcctcccc tgccccaatc
 901 ccttattac cccctccctc agacaccctc aacctttct ggctaaaaa gagaattggg
 961 ggcttagggt cggaacccaa gcttagaact ttaagcaaca agaccaccac ttgaaaccc
 1021 gggattcagg aatgtgtggc ctgcacagt aattgctggc aaccactaag aattcaaact
 1081 gggccctcca gaactactg gggctacag ctttgatccc tgacatctgg aatctggaga
 1141 ccagggagcc ttgggtctg gccagaatgc tgcaggactt gagaagacct cacctagaaa
 1201 ttgacacaag tggaccctag gcctccctc ctccagatgt ttccagactt ccttgagaca
 1261 cggagcccag ccctccccat ggagccagct ccctctattt atgtttgcac ttgtgattat
 1321 ttattattta ttattatttt attattttac agatgaatgt atttatttgg gagaccgggg
 1381 tattctgggg gaccaatgt aggagctgcc ttggctcaga catgtttcc gtgaaaacgg
 1441 agctgaacaa taggctgttc ccatgttagcc ccctggcctc tttgcctctt ttgattatg
 1501 tttttaaaaa tatttatctg attaagtgtt ctaaacaatg ctgatttggt gaccaactgt
 1561 cactcattgc tgagcctctg ctcccccagg gagttgtgtc tggatcgcc ctactattca
 1621 gtggcgagaa ataaagtttgc ctt (SEQ ID NO: 6)

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 152
- (2) 3' Untranslated Region - nts 852 - 1643

20 Initial Specific Target Motif:

Group I AU-Rich Element (ARE) Cluster in 3' untranslated region

5' AUUUAUUUAUUUAUUUAUUUA 3' (SEQ ID NO: 1)

25 6.2. Granulocyte-macrophage Colony Stimulating Factor (“GM-CSF”)
 GenBank Accession # NM_000758:

1 gctggaggat gtggctgcag agcctgctgc tcttgggcac tggccctgc agcatctgt
 61 cacccgccccg ctgcggccagc cccagcacgc agccctggga gcatgtaat gccatccagg
 121 agccccggcg ttcctgaac ctgagtagag acactgtgc tgagatgaat gaaacagtag
 181 aagtcatctc agaaatgtt gacccctcagg agccgacctg cttacagacc cgcctggagc
 241 tgtacaagca gggctgccc ggcagcctca ccaagctaa gggcccttg accatgtatgg
 30 301 ccagccacta caagcagcac tgccctccaa ccccgaaac ttccctgtca accccagacta
 361 tcaccttga aagttcaaa gagaacctga aggacttct gcttgcattc ccccttgact
 421 gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaaagcc ggggagctgc
 481 tctctcatga aacaagagct agaaactcag gatggatc ttggaggac caaggggtgg
 541 gccacagcca tggtggagttt ggcctggacc tgccctgggc cacactgacc ctgatacagg

601 catggcagaa gaatggaaat atttatact gacagaaatc agtaatattt atataatttt
 661 attttaaaa tatttattt tttatttttta taagtcata ttccatattt attcaagatg
 721 ttttaccgta ataattatta ttaaaaatat gcttct (SEQ ID NO: 7)

5

GenBank Accession # XM_003751:

1 tctggaggat gtggctgcag agcctgtgc tcttggcac tggcctgc agcatctg
 61 caccggcccg ctgcggccagc cccagcacgc agccctggga gcatgtaaat gccatccagg
 121 aggcccgccg ttcctgaac ctgagtagag acactgtgc tgagatgaat gaaacagtag
 181 aagtcatctc agaaatgtt gacccaggag agccgacccg cctacagacc cgcctggagc
 241 tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggcccttg accatgatgg
 301 ccagccacta caagcagcac tgcctccaa cccggaaac ttctgtgca acccagacta
 361 tcaccttga aagttcaaa gagaacctga aggacttct gcttgtcatc cccttgact
 421 gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc
 481 tctctcatga aacaagagct agaaactcag gatggtcac ttggagggac caaggggtgg
 541 gccacagcca tggggaggt ggcctggacc tgcctggc cacactgacc ctgatacagg
 601 catggcagaa gaatggaaat atttatact gacagaaatc agtaatattt atataatttt
 661 attttaaaa tatttattt tttattttta taagtcata ttccatattt attcaagatg
 721 ttttaccgta ataattatta ttaaaaatat gcttct (SEQ ID NO: 8)

20

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 32
- (2) 3' Untranslated Region - nts 468 - 789

25

Initial Specific Target Motif:

Group I AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' AUUUAUUUAUUUAUUUAUUA 3' (SEQ ID NO: 1)

6.3. Interleukin 2 ("IL-2")

30

GenBank Accession # U25676:

1 atcactctct ttaatcacta ctcacattaa cctcaactcc tgccacaatg tacaggatgc
 61 aactcctgtc ttgcattgca ctaattcttg cacttgtcac aaacagtgc cctacttcaa
 121 gttcgacaaa gaaaacaaag aaaacacagc tacaactgga gcatttactg ctggattac
 181 agatgattt gaatggaaatt aataattaca agaatccaa actcaccagg atgctcacat
 241 ttaagttta catgccaag aaggccacag aactgaaaca gcttcagtg ctagaagaag
 301 aactcaaacc tctggaggaa gtgctgaatt tagctcaaag caaaaactt cacttaagac

361 ccagggactt aatcagcaat atcaacgtaa tagttctgga actaaaggga tctgaaacaa
 421 cattcatgtg tgaatatgca gatgagacag caaccattgt agaatttctg aacagatgga
 481 ttacctttg tcaaagcatc atctcaacac taacttgata attaagtgtc tcccacttaa
 5 541 aacatatcag gccttctatt tatttattta aatatttaaa ttttatattt atttgtgaat
 601 gtatggttgc tacctattgt aacttattt ctaatctta aaactataaa tatggatctt
 661 ttatgattct ttttgaagc cctaggggct ctaaaatggt ttaccttattt tatcccaaaa
 721 atatttatta ttatgtgaa tgtaaatat agtatctatg tagattgggt agtaaaaacta
 781 ttaataaaat ttgataaata taaaaaaaaaa aaacaaaaaaaaaa aaaaa (SEQ ID NO: 9)

10 General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 47
- (2) 3' Untranslated Region - nts 519- 825

15 Initial Specific Target Motifs:

Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

6.4. Interleukin 6 ("IL-6")

20 GenBank Accession # NM_000600:

1 ttctgcccctc gagccccaccg ggaacgaaag agaagctcta tctgcctcc aggagccccag
 61 ctatgaactc ctttccaca agcgccctcg gtccagttgc ctttccctg gggctgcctcc
 121 tgggtttgcc tgctgccttc cctgccccag tacccccagg agaagattcc aaagatgttag
 181 ccggccccaca cagacagccca ctcacccctt cagaacgaat tgacaaacaa attcggtaca
 241 tcctcgacgg catctcagcc ctgagaaagg agacatgtaa caagagtaac atgtgtgaaa
 301 gcagcaaaga ggcactggca gaaaacaacc tgaacctcc aaagatggct gaaaaagatg
 361 gatgttcca atctggattc aatgaggaga ctggcctggt gaaaatcatc actggcttt
 421 tggagttga ggtataccta gagtacctcc agaacagatt tgagagtagt gaggaacaag
 481 ccagagctgt gcagatgagt acaaaagtcc tcatccagtt cctgcagaaaa aaggcaaaaga
 541 atctagatgc aataaccacc cctgacccaa ccacaaatgc cagcctgctg acgaagctgc
 601 aggcacagaa ccagtggctg caggacatga caactcatct cattctgcgc agcttaagg
 661 agttcctgca gtccagccctg agggcttcc ggcaatgtt gcatggcac ctcagattgt
 721 tgggtttat gggcattcc tcttctggtc agaaacctgt ccactggca cagaacttat
 781 gttgttctt atggagaact aaaagtatga gcgtaggac actatttaa ttattttaa
 841 ttatataata tttaaatatg tgaagctgag ttaattatg taagtcatat ttatatttt
 901 aagaagtacc acttggaaaca ttatgtat tagtttggaa ataataatgg aaagtggcta

961 tgcagttga atatccttg tttcagagcc agatcattc ttggaaagtg taggcttacc
 1021 tcaaataaat ggctaactta tacatatttt taaagaaata ttatattgt atttatataa
 1081 tgtataaaatg gttttatac caataaatgg catttaaaa aattc (SEQ ID NO: 11)

5

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 62
- (2) 3' Untranslated Region - nts 699 - 1125

10

Initial Specific Target Motifs:

Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

6.5. Vascular Endothelial Growth Factor ("VEGF")

15

GenBank Accession # AF022375:

20

1 aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagagc gcgcggcggt
 61 gcgagcagcg aaagcgacag gggcaaagtg agtgacctgc tttgggggt gaccgcccga
 121 gcgcggcggt agccctcccc cttggatcc cgacgtcgac cagtgcgcgt gacggacaga
 181 cagacagaca cccgccccag ccccagttac caccctctcc cccgcggcg gcccacagt
 241 gacgcggcg cgagccgcgg gcagggcccg gagccgcggcc cccggaggcg ggtggagggg
 301 gtcggagctc gcggcgctcg actgaaactt ttctgtccaa ttctggctgt ttctcgcttc
 361 ggaggagccg tggtcgcgcg ggggaaagcc gagccgagcg gagccgcgag aagtgcgtac
 421 tcggccggg aggagccgca gcggaggag ggggaggagg aagaagagaa ggaagaggag
 481 agggggccgc actggcgact cggcgctcg aagccgggt catggacggg tgaggccgc
 541 gtgtgcgcag acagtgcgtcc agcgcgcgcg ctcccgaccc ctggccggc ctggggccgg
 601 gaggaaaggt agtcgcgcg ggcgcgcgagg agagccggcc gcccacagc ccgagccgga
 661 gagggacgcg agccgcgcgc cccggcggtt cctccgaaac catgaacttt ctgctgtctt
 721 ggggcattt gggccgttgc ttgtgtctt acctccacca tgccaaatgg tccaggctg
 781 cacccatggc agaaggagga gggcagaatc atcacgaagt ggtgaagttc atggatgtct
 841 atcagcgcag ctactgcctt ccaatcgaga cccgggttga catcttcag ggttccctt
 901 atgagatcga gtacatctt aagccatctt ttgtgtccctt gatgcgttc gggggctgt
 961 ccaatgcgca gggcgtggag tttgtgtccca ctggggatgc caacatcacc atgcagatta
 1021 tgcggatcaa acctcaccaa ggccagcaca taggagagat gagcttccta cagcacaaca
 1081 aatgtgaatg cagaccaaag aaagatagag caagacaaga aaatccctgt gggccgttgc
 1141 cagagcggag aaagcatttg ttgtacaag atccgcagac gtgtaaatgt tcctgcaaaa
 1201 acacacactc gcgttgcaag ggcaggcagc ttgagttaaa cgaacgtact tgcagatgt

25

30

35

1261 acaagccgag gcgggtgagcc gggcaggagg aaggagccctc cctcagggtt tcgggaacca
 1321 gactctctc caggaagac tgatacagaa cgatcgatac agaaaccacg ctggccac
 1381 cacaccatca ccatcgacag aacagtccctt aatccagaaa cctgaaatga aggaagagga
 5 1441 gactctgcgc agagcacittt gggtccggag ggcgagactc cggcgaagc attcccgcc
 1501 gggtgaccca gcacggtccc tcttggaaatt ggattcgcca ttttatttt ctgctgta
 1561 aatcaccgag cccggaagat tagagagttt tatttctggg attcctgttag acacacccac
 1621 ccacatacat acatttatat atatatatat tatatatataaaaaataaa tatctctatt
 1681 ttatatatat aaaaatata tattttttt taaaattaac agtgctaattt ttattgggt
 10 1741 cttcactgga ttttttttttgc tgctgtggac ttgagttggg agggaaatgt tcccactcag
 1801 atcctgacag ggaagaggag gagatgagag actctggcat gatctttttt ttgtccact
 1861 tggtggggcc agggctctt cccctgccc agaatgtgca aggccaggc atgggggca
 1921 atatgaccca gttttggaa caccgacaaa cccagccctg ggcgtgagcc tcttacccc
 1981 aggtcagacg gacagaaaaga caaatcacag gttccggat gaggacacccg gctctgacca
 2041 ggagtttggg gagcttcagg acattgtgt gttttgggaa ttccctccac atgctgac
 15 2101 cgcacatctcgc ccccaaggggc actgccttggaa agattcagga gcctggcgg ccttcgttta
 2161 ctctcacctg ctctgtttt gcccaggagg ccactggcag atgtccggc gaagagaaga
 2221 gacacattgt tggaagaagc agcccatgac agcgcctt cctggactc gccctcatcc
 2281 tcttcctgct ccccttcctg gggcagcc taaaaggacc tatgtccca caccattgaa
 2341 accacttagtt ctgtccccc agggaaacctg gttgtgtgt tgtagtggt tgacccct
 20 2401 ccatccctg gtcctccct tcccttcctg aggcacagag agacaggcga ggatccacgt
 2461 gcccttgcg gaggcagaga aaagagaaaag tgttttat acggtaatcc ttaatatcc
 2521 ctttttaatt agaaattaga acagtttaatt taattaaaga gttaggtttt ttttcagtt
 2581 tcttggtaa tatttaattt caactattt tgagatgtat ctttgctct ctctgctct
 2641 ctttttgta cccgttttg tatataaaat tcatgtttcc aatctctc tccctgatcg
 2701 gtgacagtca ctatgttatac ttgaacagat atttaattt gctaacactc agctctgccc
 2761 tcccccgttcc cctggctcc cagcacacat tcctttgaaa gagggttca atatacatct
 2821 acatactata tatataattgg gcaacttgta ttgtgtgtatata tatatgtta
 2881 tgtatataatg tgatcctgaa aaaataaaca tcgctattct gtttttata tggtcaaaacc
 2941 aaacaagaaaa aaatagagaaa ttctacatac taaaatctc tccctttta atttaatatt
 30 3001 ttgttatcat ttattttatg gtgctactgt ttatccgtaa taattgtggg gaaaagatatt
 3061 taacatcactg tcttgcgtt tagtgcagtt ttgcagata ttccgttagta catatttatt
 3121 tttaaaacaac gacaaagaaa tacagatata tcttaaaaaaa aaaaaa (SEQ ID NO: 12)

35 General Target Regions:

(1) 5' Untranslated Region - nts 1 - 701

(2) 3' Untranslated Region - nts 1275 - 3166

Initial Specific Target Motifs:

5 (1) Internal Ribosome Entry Site (IRES) in 5' untranslated region nts 513 -704
 5'CCGGGCUCAUGGACGGGUGAGGCAGCGUGUGCGCAGACAGU
 GCUCCAGCGCGCGCUCCCCAGCCCUGGCCGGCCUCGGGCCG
 GGAGGAAGAGUAGCUCGCCAGGGGCCGAGGAGAGCGGGCCGC
 CCCACAGCCGAGCCGGAGAGGGACCGCAGCCGCGCCCCGGU
 CGGGCCUCCGAAACCAUGAACUUUCUGCUGCUUGGGUGCAIU
 10 GGAGCCUUGCUCUGCUCUACCUCCACCAUG 3' (SEQ ID NO:
 13)

(2) Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

15

6.6. Human Immunodeficiency Virus I ("HIV-1")

GenBank Accession # NC_001802:

1 ggtctctctg gtttagaccag atctgagccct gggagctctc tggctaacta gggAACCCAC
 61 tgcttaagcc tcaataaaagc ttgccttgag tgcttaaggt agtgtgtgcc cgtctgttgt
 121 gtgactctgg taactagaga tccctcagac ccttttagtc agtgtggaaa atctctagca
 181 gtggcgcccg aacagggacc tgaaagcgaa agggaaacca gaggagctct ctcgacgcag
 241 gactcggctt gctgaaggcgc gcacggcaag aggcgagggg cggcgactgg tgagtacgcc
 301 aaaaaatttg actagcggag gctagaagga gagagatggg tgcgagagcg tcagtattaa
 361 gcgggggaga attagatcga tggaaaaaaa ttcggttaag gccaggggaa aagaaaaaat
 421 ataaatttaaa acatatacgta tgggcaagca gggagctaga acgattcgcgtaatccctg
 481 gcctgttaga aacatcagaa ggctgttagac aaatactggg acagctacaa ccattccctc
 541 agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tattgtgtgc
 601 atcaaaggat agagataaaaa gacaccaagg aagctttaga caagatagag gaagagcaaa
 661 acaaaagtaa gaaaaaagca cagcaagcag cagctgacac aggacacagc aatcaggta
 721 gccaaaattta ccctatagt cagaacatcc aggggcaat ggtacatcag gccatatcac
 781 ctagaacttt aaatgcattgg gtaaaaagtag tagaagagaa ggcttcagc ccagaagtga
 841 tacccatgtt ttccatgtt tcagaaggag ccaccccaaa agatttaaac accatgttaa
 901 acacagtggg gggacatcaa gcagccatgc aaatgttaaa agagaccatc aatgagggaaag
 961 ctgcagaatg ggatagagtg catccagtgc atgcagggcc tattgcacca ggcagatga
 1021 gagaaccaag gggaaagtgc atagcaggaa ctactagtac ctttcaggaa caaataggat
 1081 ggatgacaaa taatccaccc atccctgtt gggaaattta taaaagatgg ataatccctgg

1141 gattaaataa aatagtaaga atgtatagcc ctaccagcat tctggacata agacaaggac
1201 caaaggaacc ctttagagac tatgttagacc ggttctataa aactctaaga gccgagcaag
1261 ctccacagga ggtaaaaaat tggatgacag aaaccttggt ggtccaaaat gcgaaccagg
1321 attgttaagac tattttaaaaa gcattggac cagcggctac actagaagaa atgtatgacag
1381 catgtcaggg agtaggagga cccggccata aggcaagagt ttggctgaa gcaatgagcc
1441 aagtaacaaa ttcagctacc ataatgtgc agagaggcaa ttttaggaac caaagaaaaga
1501 ttgttaagtgt ttcaattgt ggcaaaagaag ggcacacagc cagaaattgc agggccctta
1561 gggaaaaggg ctgttggaaa tggaaagg aaggacacca aatgaaagat tggactgaga
1621 gacaggctaa ttttttaggg aagatctggc cttcctacaa gggaggcga gggattttc
1681 ttcagagcag accagagccaa acagccccac cagaagagag cttcaggctt ggggttagaga
1741 caacaactcc ccctcagaag caggagccga tagacaagga actgtatcc ttaacttccc
1801 tcaggtcact ctttggcaac gacccctcgta cacaataaag ataggggggc aactaaaggaa
1861 agctctatta gatacaggag cagatgatac agtatttagaa gaaatgagggt tgccaggaag
1921 atggaaacca aaaatgtatgggggaaattgg aggtttatc aaagtaagac agtattgtatca
1981 gatactcata gaaatctgtg gacataaagc tatagttaca gtatttagtag gacctacacc
2041 tgtcaacata attggaagaa atctgttgac tcagattggt tgcaatttaa attttcccat
2101 tagccctatt gagactgtac cagtaaaattt aaagccagga atggatggcc caaaagttaa
2161 acaatggcca ttgacagaag aaaaaataaa agcatttagta gaaatttgta cagagatgga
2221 aaaggaaaggaa aaaaatttcaa aaattggcc tgaaaatcca tacaataactc cagtattgc
2281 catabagaaa aaagacagta ctaaatggag aaaaatttagta gatttcagag aacttaataa
2341 gagaactcaa gacttctggg aagtcaattt aggaataccaa catccgcag ggtttaaaaaaa
2401 gaaaaaaatca gtaacagtac tggatgtggg tggatgtat tttcagttc ctttagatga
2461 agacttcagg aagtatactg catttaccat accttagtata aacaatgaga caccaggat
2521 tagatatcag tacaatgtgc ttccacaggg atggaaagga tcaccagcaa tattccaaag
2581 tagcatgaca aaaatcttag agccttttag aaaaacaaaat ccagacatag ttatctatca
2641 atacatggat gattttatg taggatctga ctttagaaata gggcagcata gaacaaaaat
2701 agaggagctg agacaacatc tggatgtggg gggacttacc acaccagaca aaaaacatca
2761 gaaagaacct ccattccctt ggtatgggta tgaactccat cctgataaat ggacagtaca
2821 gcctatagtg ctggcagaaa aagacagctg gactgtcaat gacatacaga agttagtggg
2881 gaaattgaat tggcaagtc agatttaccc agggattaaa gtaaggcaat tatgtaaact
2941 ctttagagga accaaagcac taacagaagt aataccacta acagaagaag cagagctaga
3001 actggcagaa aacagagaga ttctaaaaga accagtacat ggagtgtatt atgacccatc
3061 aaaagactta atagcagaaa tacagaagca gggcaaggc caatggacat atcaaattta
3121 tcaagagcca tttaaaaatc tgaaaacagg aaaatatgca agaatgaggg gtgccacac
3181 taatgtatgtaa aacaattaa cagaggcagt gcaaaaaata accacagaaa gcatagtaat

3241 atggggaaag actcctaaat taaaactgcc cataaaaaag gaaacatggg aaacatggg
3301 gacagagtat tggcaagcca cctggattcc tgagtggag ttgttaata cccctccctt
3361 agtggaaatta tggtaccagt tagagaaaga acccatagta ggagcagaaa cctctatgt
3421 agatggggca gctaacaggg agactaaatt aggaaaagca ggataatgtta ctaatagagg
3481 aagacaaaaaa gttgtcaccc taactgacac aacaaatcg aagactgagt tacaagcaat
3541 ttatctagct ttgcaggatt cgggattaga agtaaacata gtaacagact cacaatatgc
3601 attaggaatc attcaaggcac aaccagatca aagtgaatca gagttatgtca atcaaataat
3661 agagcgtta ataaaaaagg aaaaggtcta tctggatgg gtaccagcac acaaaggaat
3721 tggagggaaat gaacaagtag ataaatttagt cagtgtcgga atcagggaaag tactatttt
3781 agatggaaata gataaggccc aagatgaaca tgagaaatatacagtaatt ggagagcaat
3841 ggctgtgtat ttacacccgc cacctgttagt agcaaaaagaa atagtagccca gctgtgataa
3901 atgtcagcta aaaggagaag ccatgcatgg acaagtagac ttagtccag gaatatggca
3961 actagattgt acacatttag aaggaaaagt tatcctggta gcagttcatg tagccagtgg
4021 atatatagaa gcagaagtttta ccagcaga aacagggcag gaaacagcat attttttt
4081 aaaatttagca ggaagatggc cagtaaaaac aatacatact gacaatggca gcaatttcac
4141 cggtgctacg gttagggccg cctgtgggt ggcgggaatc aagcaggaat ttggaaattcc
4201 ctacaatccc caaagtcaag gtagtagta atctatgaat aaagaattaa agaaaaattat
4261 aggacaggtta agagatcagg ctgaacatct taagacagca gtacaaatgg cagtattcat
4321 ccacaattttt aaaagaaaaag gggggattgg ggggtacagt gcaggggaaa gaatagttaga
4381 cataatagca acagacatac aaactaaaga attacaaaaaa caaattacaa aaattcaaaa
4441 tttcgggtt tattacaggg acagcagaaa tccactttgg aaaggaccag caaagtcct
4501 ctggaaaggt gaagggcag tagtaataca agataatagt gacataaaaag tagtccaaag
4561 aagaaaaagca aagatcatta gggattatgg aaaacagatg gcaggtgtat attgtgtggc
4621 aagtagacag gatgaggatt agaacatggaa aaagttagt aaaacaccat atgtatgtt
4681 cagggaaagc tagggatgg tttatagac atcactatga aagccctcat ccaagaataa
4741 gttcagaagt acacatccca ctagggatg ctagatttgtt aataacaaca tattggggtc
4801 tgcatacagg agaaagagac tggcatttg gtcagggagt ctccatagaa tggagggaaa
4861 agagatatac cacacaagta gaccctgaac tagcagacca actaattcat ctgtattact
4921 ttgactgtttt ttcagactct gctataagaa aggcccttattt aggacacata gttagcccta
4981 ggtgtgaata tcaagcagga cataacaagg taggatctct acaatacttg gcactagcag
5041 cattaataac accaaaaaaag ataaagccac cttgccttag tggatcgaaa ctgacagagg
5101 atagatggaa caagccccag aagaccaagg gcccacaggg gggccacaca atgaatggac
5161 actagagctt ttagaggagc ttaagaatga agctgttaga cattttccctt ggattttggct
5221 ccatggctta gggcaacata tctatgaaac ttatggggat actggcag ggtgtggaaagc
5281 cataataaga attctgcac aactgtgtt tatccatttt cagaattggg tggatcgacata

7441 ggctatttag ggcacacagg atctgttgc actcacagtc tggggcatca agcagctcca
7501 ggcagaatc ctggctgtgg aaagataacct aaaggatcaa cagctctgg ggattttgggg
7561 ttgctctgga aaactcattt gcaccactgc tgtgccttgg aatgcttagtt ggagtaataa
5 7621 atctctggaa cagatttggaa atcacacgac ctggatggag tggacagagaaatcaa
7681 ttacacaagg ttaatacact ccttaatttga agaatcgcaa aaccagcaag aaaagaatga
7741 acaagaatta ttgaaatttag ataaatggc aagtttgggg aattgggtta acataacaaa
7801 ttgctgtgg tatataaaat tattcataat gatagtagga ggcttggtag gttttagaaat
7861 agttttgtct gtactttcta tagtgaatag agtttaggcag ggatattcac cattatcggt
10 7921 tcagacccac ctcccaaccc cgaggggacc cgacaggccc gaaggaatag aagaagaagg
7981 tggagagaga gacagagaca gatccattcg attagtgaac ggatccttgg cacttatctg
8041 ggacgatctg cggagcctgt gccttctcag ctaccaccgc ttgagagact tactcttgat
8101 tggtaacgggg atttgggaaac ttctgggacg caggggggggg gaagccctca aatattggtg
8161 gaatctccta cagtatttggaa gtcaggaact aaagaatagt gctgttagt tgctcaatgc
15 8221 cacagccata gcagtagctg agggggacaga taggggtata gaagtagtac aaggagcttg
8281 tagagctattt cggccacatac cttagaagaat aagacaggc ttggaaaggg ttttgcata
8341 agatgggtgg caagtggtca aaaagtagtg tgattggatg gcctactgt aaggaaagaa
8401 tgagacgagc tgagccagca gcagataggg tggggcagc atctcgagac ctggaaaaac
8461 atggagcaat cacaagtagc aatacagcag ctaccaatgc tgcttgcc tggctagaag
20 8521 cacaagagga ggaggaggtg ggtttccag tcacaccctca ggtaccccttia agaccaatga
8581 cttacaaggc agctgttagat ctttagccact ttttaaaaga aaaggggggg ctggaaaggc
8641 taatttcaactc ccaaagaaga caagatatcc ttgatctgt gatctaccac acacaaggct
8701 acttccctga tttagcagaac tacacaccag ggccagggggt cagatatcca ctgaccccttg
8761 gatggtgcta caagcttagta ccagttgagc cagataagat agaagaggcc aataaaggag
25 8821 agaacaccag ctgttacac cctgtgagcc tgcattggat ggatgaccgg gagagagaag
8881 tggtagatg gagggttgc acggccctag catttcatca cgtggcccg gagctgcac
8941 cggagtagtctt caagaactgc tgacatcgag cttgctacaa gggacttcc gctggggact
9001 ttccaggggag gctggccctg ggcgggactg gggagttggcg agccctcaga tcctgcata
9061 aagcagctgc ttttgcctg tactgggtct ctctggtag accagatctg agcctgggag
30 9121 ctctctggct aactaggaa cccactgtt aagcctcaat aaagcttgcc ttgagtgctt
9181 c (SEQ ID NO: 14)

Initial Specific Target Motifs:

(1) Trans-activation response region/Tat protein binding site - TAR RNA - nts 1
35 - 60
"Minimal" TAR RNA element

5' GGCAGAUCUGAGCCUGGGAGCUCUCUGCC 3' (SEQ ID NO: 15)

(2) Gag/Pol Frameshifting Site - "Minimal" frameshifting element

5'

5 UUUUUUAGGGAAGAUCUGGCCUUCUACAAGGGAAGGCCAGG
GAAUUUUCUU 3' (SEQ ID NO: 16)

6.7. Hepatitis C Virus ("HCV" - Genotypes 1a & 1b)

GenBank Accession # NC_001433:

10 1 ttggggcga cactccacca tagatcactc ccctgtgagg aactactgtc ttcacgcaga
61 aagegtctag ccatggcggtt agtatgatgt ttgtgcagcc tccaggaccc cccctcccg
121 gagagccata gtggctcgcg gaaccggta gtacaccgga attgccagga cgaccgggtc
181 ctttcttggaa tcaaccgcgtt caatgcctgg agatttgggc gtgcggccgc gagactgcta
241 gcccggatgtt gttgggtcgc gaaaggcctt gtggtaactgc ctgtatagggt gcttgcgagt
301 gccccgggag gtctcgtaga ccgtgcacca tgagcacaaa tcctaaacct caaagaaaaa
361 ccaaactgtaa caccaaccgc cgcccacagg acgttaagtt cccggcggtt ggtcagatcg
421 ttggtggagt ttacctgttg ccgcgcagggg gcccagggtt ggggtgtgcgc gcgacttagga
481 agacttccga gcggtcgcaa cctcggtggaa ggcgacaacc tatcccaag gtcgcggc
541 ccgagggttag gacctgggtt cagccgggtt acccttggcc cctctatggc aacgagggtt
601 tgggtgggc aggtggctc ctgtcaccc ttgtgcgttgc gcctagttgg ggccccacag
661 acccccggcg taggtcgctt aatttgggtt aggtcatcgta tacccttaca tgcgggttcg
721 ccgacccatggggtacatt ccgcgtgtcg ggcggccctt agggggcgctt gccaggccc
781 tggcacatgg tgcgggtt ctggaggacg gcgtgaacta tgcaacaggg aatctggcc
841 gttgtcttt ctctatcttc ctcttagctt tgctgtcttg ttgtaccatc ccagcttccg
901 cttagggatgtt ggcacacgtt tccggatattt accatgtcac gaacgactgc tccaaactcaa
961 gtattgtgtt tgaggcagcg gacatgtatca tgcacaccc cgggtgcgtt ccctgcgtcc
1021 gggagagtaa ttctccgtt tgctgggtt cgttcactcc cacgtcgcc gcaaggaaaca
1081 gcagcatccc caccacgaca atacgacgcc acgtcgattt gctcggtgg gcccgtgtc
1141 tctgtccgc tatgtacgtt gggatctctt gggatccgtt tttctcgcc tcccaactgt
1201 tcacccatcc acctcgccgg tatgagacgg tacaagattt caattgtca atctatcccg
1261 gcccacgtatc aggttcaccgc atggcttggg atatgtatgtt gaaactgggtca cctacaacgg
1321 cccttaggtt atcgacgtt ccgcgttgcctt cacaaggccgtt cgtggacatgtt gtggcgcccc
1381 cccactgggg tgccttagcg ggccttgcctt actattccat ggtggggaaac tgggtcaagg
1441 tttgtatgtt gatgtacttc ttgtgttgcgtt tgacgggcata caccacgtt acagggggaa
1501 gggtagcctc cagcacccag agcctcggtt cctggctctc acaaggccca tctcagaaaaa
1561 tccaaactcgtaa gacaccaac ggcagctggc acatcaacag gaccgtctgtt aattgcaatg

1621 actccctcca aactgggttc attgctgcgc tggctacgc acacagggtc aacgcgtccg
1681 ggtgccaga ggcgcgttgc agctggccgc ccatcgatga gttcgctcag gggggggc
1741 ccatcactca tggatatgcct gagagctcg accagaggcc atattgctgg cactacgcgc
5 1801 ctcgaccgtg cgggatcggt cctgcgtcgc aggtgtgtgg tccagtgtat tgcttactc
1861 cgagccctgt tggatgggg acgaccgatc gttcgccgc tcctacgtat agctgggggg
1921 agaatgagac agacgtgctg ctacttagca acacgcggcc gcctcaaggc aactgggttg
1981 ggtgcacgtg gatgaacgcg actgggttca ccaagacgtg cggggccct ccgtgcaaca
2041 tcgggggggt cggcaacaac acctgggtct gccccacggg ttgctccgg aagcaccccg
10 2101 aggccactta cacaaggatgt ggctggggc cctgggtgac acccaggtgc atgggtgact
2161 acccatacag gctctggcac taccctgca ctgttaactt taccgtctt aaggtcagga
2221 tggatgtggg gggcgtggag cacaggctca atgctgcgtg caattggact cgaggagagc
2281 gctgtgactt ggaggacagg gataggtcag aactcagccc gctgtgctg tctacaacag
2341 atggccagat actgcctgt tccttcacca ccctaccggc cctgtccact ggcttgatcc
15 2401 atcttcacccg gaacatcggt gacgtgcaat acctgtacgg tatagggtcg gcagggtct
2461 ccttgcataat caaatgggag tatatcctgt tgctttctt tcttcggcg gacgcgcgcg
2521 tctgtgcctg cttgtggatg atgctgcgtg tagcccaggc tgaggccacc tttagagaacc
2581 tggggccctt caatgcggcg tctgtggccg gagcgcgttgg ccttcctcc ttctcgtgt
2641 tcttcgcgc cgcctggatc atcaaaggca ggctggccc tggggccgc tatgtctct
20 2701 atggcgtatg ggcgttgcctc ctgccttgc tggccttacc accacgagct tatgcctgg
2761 accgagagat ggctgcgtg tgcggaggcg cgggttttgtt aggtctggta ctcttgcac
2821 tggatgtgggatgttccctgcata ggctcatatg tgggttacaa tattttatca
2881 ccagagccga ggcgcacttg caagtgtggg tccccctct caatgttcgg ggaggccgc
2941 atgcctatcat ctccttaca tgcgcggcctt atccagagct aatcttgc atcaccaaac
25 3001 tcctgctgc catactcggt ccgcgtatgg tgcctcaggc tggcataact agagtgcgc
3061 actttgtacg cgctcagggg ctcatcggt catgcgtt agtgcggaaag gtcgtggag
3121 ggcactatgt ccaaattggcc ttcatgaagc tggccgcgt gacaggtacg tacgtatatg
3181 accatcttac tccactgcgg gattggggcc acgcggccctt acgagaccctt ggcgtggcag
3241 tagagcccgatgttccctgcata gacatggaga cttaactcat cacctggggg gcagacaccg
30 3301 cggcgtgtgg ggacatcatc tgggtctac cagtcctccgc ccgaaggggg aaggagatac
3361 ttcttaggacc ggcgcataatgt tttggagacg aggggtggcg gtccttgcgc cctatcaccgg
3421 cctatccca cacaacgcgg ggcctgttgc gctgtatcat cactagccctc acaggcgggg
3481 acaagaacca ggtcgatggg gagggtcagg tgcttccac cgcaacgcaa tctttctgg
3541 cgacatgcgtt cttatggcgatgt tggatggaccg tctaccatgg tgccggctcg aagaccctgg
35 3601 ccggcccgaa gggtccaatc acccaaattgtt acaccaatgtt agaccaggac ctcgtggct
3661 ggcggccgc cccggggccgatgtt cttatggcgatgtt cttatggcgatgtt tggaccctt

5 5821 gcgccggcat tgccggtgcg gccgtggca gcataggctc cgggaaggta ctttgtggaca
5881 ttctggcggg ctatggggcg ggggtggctg ggcactcg ggccttaag gtcgtggacg
5941 gcgagatgcc ctccactgag gatctggta attactccc tgccatcct tctctggcg
6001 ccctgggtgt cgggtcggt tgccgaccaa tactgcgtcg gcacgtggc cggggagagg
6061 gggctgtgca gtggatgaac cggctgatag cgttcgttc gcggggtaac cacgtctccc
6121 ccacgcacta tgtccccag agcgacgccc cggcgcgtgt tactcagatc ctctccagcc
6181 ttaccatcac tcagttgctg aagaggcttc atcagtggat taatgaggac tgctccacgc
6241 ctgttccgg ctcgtggcta aaggatgtt gggactggat atgcacgggt tgtagtact
10 6301 tcaagacttg gctccagtc aagctcctgc cgcggttacc gggactccct ttccgtcat
6361 gccaacgegg gtacaaggga gtctggcggg gggatggcat catgcaaacc acctgcccatt
6421 gtggagcaca gatcaccggaa catgtaaaa atggctcat gaggattgtt gggccaaaa
6481 cctgcagcaa cacgtggcat ggaacattcc ccatcaacgc atacaccacg gccccctgca
6541 cggccctcccc agcgccgaac tattccaggg cgctgtggcg ggtggctgt gaggagtacg
6601 tggaggttac gcggtgggg gatttccact acgtgacggg catgaccact gacaacgtga
6661 aatgcccatttccatg ccaggttccaa gcccctgaat tttcacggaa ggtggatggaa gtacgggttgc
6721 acaggtatgc tccagtgtgc aaacctctcc tacgagagga ggtcgtattc caggtggc
6781 tcaaccagta cctggcgaaa tcacagtc catgtgagcc cgaaccggat gtggcagtgc
6841 tcacttccat gtcaccgac ccctctcata ttacagcaga gacggccaag cgtaggctgg
20 6901 ccaggggttc tccccctcc ttggccagct cttagctag ccagttgtc ggccttctt
6961 tgaaggcgcac atgtactacc catcatgact cccggacgc tgacctcatc gaggccaaacc
7021 tcctgtggcg gcaggagatg ggcgggaaca tcacccgtgt ggagtcagaa aataaggtgg
7081 taatcctggaa ctttcgat ccgattcgaaa cggtgagggaa tgagagggaa atatccgtcc
7141 cggccggagat cctgcgaaaa cccaggaagt tccccccagc gttgcccata tgggcacgc
25 7201 cggattacaa ccctccactg cttagtgc tggacggaccc ggactacgtc ccccccgtgg
7261 tacacgggtg cccttgcata tctaccaagg ccccccaat accacccatca cggagggaaa
7321 ggacgggtgt cctgacagag tccaccgtgt ctctgcctt ggcggagctc gctactaaga
7381 cccttggcag ctccgggtcg tcggccgttg acagcggcac ggcgacttgc ccccccgtac
7441 aggctccga cgacggcgac aaaggatccg acgttgagtc gtactccctt atgccccccc
7501 tcgagggaga gccaggggac cccgacccca ggcacgggtc ttggtctacc gtggcgggg
30 7561 aagctggtga ggacgtcgac tgctgctcaa tgcctatac atggacaggt gcttgcata
7621 cgccatgcgc tgccggaggag agcaagttgc ccatcaatcc gttgagcaac tctttgtgc
7681 gtcaccacag tatggtctac tccacaacat ctgcagcgc aagtctgcgg cagaagaagg
7741 tcacccatgaa cagactgcaaa gtcctggacg accactaccg ggacgtgtc aaggagatga
7801 aggccaaaggc gtccacagtt aaggcttaggc ttctatctat agaggaggcc tgcaactgaa
35 7861 cggcccccaca ttccggccaaa tccaaatttg gtcacggggc gaaggacgtc cggggctat

7921 ccagcaggc cgtaaccac atccgctccg tggggagga ctgcgtgaa gacactgaaa
7981 caccaattga taccaccatc atggcaaaaa atgagggttt ctgcgtccaa ccagagaaag
8041 gaggccgcaa gccagctcgc citatcgat tcccagacct ggggtacgt gtatcgaga
8101 agatggccct ttacgacgtg gtctccaccc ttccctcaggc cgtatggc ccctcatacg
8161 gattccagta ctctctggg cagcgggtcg agttcctggt gaataccctgg aatcaaaga
8221 aatgccatat gggcttctca tatgacaccc gctgcttga ctcaacggc actgagaatg
8281 acatccgtac tgaggaatca atttaccaat gttgtgactt ggcccccga gccaggcagg
8341 ccataaggc gctcacagag cggctttag tcgggggtcc cctgactaat tcgaaggggc
8401 agaactgcgg ttatcgccgg tgccgcgcaa gtggcgtct gacgactagc tgccgcaca
8461 ccctcacatg ttacttgaag gccactgcgg cctgtcgagc tgcaaagctc cagggactgca
8521 cgtatcgctgt gaacggagac gaccctgtcg ttatctgtga gagtgcggga acccaggagg
8581 atgcggcggc cctacgagcc ttacggagg ctatgacttag gtattccgccc cccccgggg
8641 acccgccccca accagaatac gacttggagc tgataacgac atgctctcc aatgtgtcgg
8701 tcgcgcacga tgcacccggc aaaagggtgt actacccac ccgtgacccc accacccccc
8761 tcgcacgggc tgctggag acagtttagac acactccagt caactcctgg ctggcaata
8821 tcatcatgtt tgcccccacc ctatggcga ggatgattct gatgactcat ttctctcta
8881 tccttcttagc tcaaggagcaa ctggaaaaag ccctggattt tcaatctac ggggcgtt
8941 actccattga gccacttgcac ctacctcaga tcattgaacg actccatgtt cttagcgcac
9001 ttcaactcca cagttactct ccaggtgaga tcaatagggt ggctcatgc ctcaggaaac
9061 ttggggtacc gccttgcga gtctggagac atcggccag aagtgtccgc gctaagctac
9121 tgtcccaggg ggggagggtt gccacttgcg gcaagtacctt ctcaactgg gcagtaaaga
9181 ccaagcttaa actcaactcca atccggctcg cgtcccagt agacttgcg ggctgggtcg
9241 ttgtctggta caacggggga gacatatac acagcctgc tcgtcccgaa cccctgg
9301 tcaatgttgc cttactccta cttctgttag gggtaggcat ctacccgtc cccaaacccgt
9361 gaacggggag ctaaccactc caggccaaata ggccattccc ttttttttt ttc (SEQ ID NO: 17)

General Target Region:

5' Untranslated Region - nts 1 - 328 - Internal Ribosome Entry Site (IRES):

30 5'UUGGGGGCGACACUCCACCAUAGAUCACUCUCCCCUGUGAGGAACUACUGUCU
UCACGCAGAAAGCGUCUAGCCAUGGCGUUAGUAUGAGUGUUGUGCAGCCUC
CAGGACCCCCCCCUCCCGGGAGAGCCAUAGUGGUCUGCGAACCGGUGAGUAC
ACCGGAAUUGCCAGGACGACCGGGGUCCUUUCUUGGAUCAACCCGCUCAAUGC
CUGGAGAUUUGGGCGUGCCCCCGCGAGACUGCUAGCCGAGUAGUGUUGGGU
35 CGCGAAAGGCCUUGUGGUACUGCCUGAUAGGGUGCUUGCGAGUGCCCCGGG
AGGUCUCGUAGACCGUGCAU3' (SEQ ID NO: 18)

Initial Specific Target Motifs:

5 (1) Subdomain IIIc within HCV IRES - nts 213 - 226
5'AUUUUGGGCGUGGCC3' (SEQ ID NO: 19)
(2) Subdomain IIId within HCV IRES - nts 241-267
5'GCCGAGUAGUGUUGGGUCGCGAAAGGC3' (SEQ ID NO: 20)

6.8. Ribonuclease P RNA (“RNaseP”)

10 GenBank Accession #s

X15624 Homo sapiens RNaseP H1 RNA:

1 atggcgagg ggaagctcat cagtggggcc acgagctgag tgcgtccctgt cactccactc
61 ccatgtccct tggaaaggc tgagactagg gccagaggcg gccctaacag ggctctccct
121 gagcttcagg gaggtgagtt cccagagaac ggggctccgc gcgaggtcag actggcagg
181 agatgccgtg gaccccgccc ttccggggagg ggcccgccgg atgcctccct tgccggagct
241 tggaacacagac tcacggccag cgaagtgagtt tcaatggctg aggtgaggta ccccgccagg
301 gacccataaa cccaaattcag accactctcc tccggccatt (SEQ ID NO: 21)

U64885 *Staphylococcus aureus* RNaseP (rrnB) RNA:

20 1 gaggaaagtc cgggctcaca cagtctgaga tgattgtagt gttcgtgtttt gatgaaacaa
61 taaatcaagg cattaatttg acggcaatga aatatcctaa gtcttcgtatggatagag
121 taatttgaaa gtgccacagt gacgttagttt atatagaaat ataaaaggtg gaacgcggta
181 aaccctcgat gtgagcaatc caaatttggt aggagcactt gtttaacggattcaacgta
241 taaacgagac acacttcgatc aatgaagtgtgttagacag atggttatcatcctgatacc
301 agtgtgacta gtgcacgtatc tgagtacgtatc ggaacagaac gcggcttat (SEQ ID NO: 22)

M17569 Escherichia coli RNA component (M1 RNA) of ribonuclease P
(rnpB) gene:
1 gaagctgacc agacagtgcg cgcttcgtcg tcgtcccttt cgggggagac gggcgagggg
30 61 gagaaaaagtc cgggctccat agggcagggt gccaggttaac gcctgggggg gaaacccacg
121 accagtgc当地 cagagagcaa accgc当地atg gccgc当地caaa gc当地ggatc当地 gtaagggt当地
181 aagggtgc当地 ggtaagagcgca cc当地gc当地ggct ggtaacagtc cgtggcac当地 taaactccac
241 cc当地ggagcaag gc当地caaatagg gg当地ttc当地ataag gtac当地gg当地cccg tactgaaccc gggtaggct当地
301 ct当地tggagccag tgagcgattt cttgc当地ctt当地aga tgaatgactt cccacgacag aacc当地ccggctt
35 361 atc当地gggtc当地tgg (SEQ ID NO: 23)

Z70692 Mycobacterium tuberculosis RNaseP (rnpB) RNA:

1 ccacccggta cgatcttgcc gaccatggcc ccacaatagg gccggggaga cccggcgtca
61 gtgggtggcg gcacggtcag taacgtctgc gcaacacggg gttgactgac gggcaatatc
5 121 ggctccatag cgtcgccgcg ggatacaga aaggagcatt ctgtacgga aaagacccc
181 gacgacgtct tcaaacttgc caaggacgag aaggtcaat atgtcgtcgt ccgggtctgt
241 gacctgcctg gcatcatgca gcacttcacg attccggctt cggcctttga caagagcgtg
301 ttggacgacg gcttggcctt tgacggctcg tcgattcgcg ggtccagtc gatccacgaa
361 tccgacatgt tgcttctcc cgatcccgag acggcgcgca tcgaccgtt ccgcgcggcc
421 aagacgctga atatcaactt tttgtgcac gacccgttca ccctggagcc gtactccgc
10 481 gacccgcgca acatcgcccg caaggccgag aactaccgtca tcagcactgg catgcgcgac
541 accgcatact tcggcgccga ggccgagttc tacattttcg attcggtgag ctgcactcg
601 cgcgcacaacg gtccttcta cgaggtggac gccatctcg ggtggtggaa caccggcgcg
661 gcgaccgagg ccgacggcag tcccaaccgg ggctacaagg tccgcaccaa gggcgggtat
15 721 ttcccaagtgg cccccaacga ccaatacgtc gacctgcgcg acaagatgtt gaccaacctg
781 atcaactccg gtttcatctt ggagaaggcc caccacgagg tggcagcgg cggacaggcc
841 gagatcaactt accagttcaa ttgcgtctg cacgcccgg acgacatgca gttgtacaag
901 tacatcatca agaacaccgc ctggcagaac ggcaaaacgg tcacgttcat gccaagccg
961 ctgttcggcg acaacgggtc cggcatgcac tgcgtcaatg cgtgtggaa ggacggggcc
20 1021 ccgctgtatgt acgacgagac gggttatgcc ggtctgtcg acacggcccg tcattacatc
1081 ggcggccctgt tacaccacgc gccgtcgctg ctggccctca ccaacccgac ggtgaactcc
1141 tacaagcggc ttgttcccg ttacgaggcc cggatcaacc ttgtctatag ccagcgcac
1201 cggtcggcat gctgtgcgtat cccgatcacc ggcagcaacc cgaaggccaa gcggctggag
1261 ttccgaagcc ccgactcgtc gggcaacccg tatctggctg tctcgccat gtcgtatggca
1321 ggcctggacg gatcaagaa caagatcgag ccgcaggcgc ccgtcgacaa ggatctctac
1381 gagctgcgcg cggaaagggc cgcgagttac ccgcagactc cgacccagct gtcagatgt
1441 atgcaccgtc tggaggccga ccacgaatac ctcacccaaag gagggtgtt cacaacgcac
1501 ctgtatcgaga cgtggatcag ttcaagcgc gaaaacgaga tcgagccgtt caacatccgg
1561 ccgcattccct acgaattcgc gctgtactac gacgttaag gacttgcg agtccgggt
30 1621 tagagggagc ggcgtgtcg tggccaggccg ggcgtcgagg ttttcgtat ggtgacgggt
1681 gcccggcaacg gcgccgcac caccgtcgca aagagccgt ttaagaacgt tcaaggacgt
1741 ttccagccggg tgccacaacc cgcgtggcaaa tcatctcccg accgcggcgc ggggtgtt
1801 tcacatgcgc cggaaactcaa gccacgtcg tgcggccgcg tgcgtcgcc gccgggtcag
1861 gttaaatgttc ggggattcgt cgtgcggccg ggcgtccacg ctgaccaacg gggcagtcaa
35 1921 ctcccaaca ctttgcgcac taccgcctt gcccggccgc tcacccgtat gtagttgtcc
1981 aggaattcccc caccgtcgccag gtttgcgcag cggccgcgaa ccgcgaccgc attgagctgg

2041 cgcccggtc cggcagctg gtcggggc ttgcgcgc ccaacaccag cgcgttgcgg
2101 gccccgggtgg cggtcagcca ggcctgacgg agcagctcca cgtggctgc gggaccaga
2161 tcggcggccg cgatgacatc cagggattc agcgtcgagg tgggtgcag ggccggaaacc
2221 tggtgccat gctgttagtc cagcaactgc acggccatt cgtgtccgc cagtcgccc
2281 cggcccagtt tgggtgtgt gttggggcgc gcaccgcgc gcaaccgctc ggactcgata
2341 cggcccttga tgcggcgaat ctgcgcacc gactcaggc acacaccgtc gggcggatac
2401 cgcgtttgt cgaccatccg taggaatcgc tgaccact cggcatcgcc ggcaaccgac
2461 tggcgcgtc gacggccctg gatctccat ggctgtccc actgctgtc gtatgcggcg
2521 taggacccca gggcgcggac cagcggaccc ttgcggccct cgggtcgaa attggcgtcg
2581 agctccagcg gcccgtcgac gctgggtgc cccagcagcg cccgaacccg ctggcgatc
2641 gatgtcgacc atttcaccgc ccgtgcacgc tcgacgcgg tggccggctc acagacgaac
2701 atcacgtcgg catccgaccc gtagccaaac tcggcaccac ccagccgacc catgccgtatc
2761 acccgatgg ccgcggggc gcgatcgac tcgggaaggc tggccggat catgacgtcc
2821 agcgcggcct gcagcacccgc cacccacacc gacgtcaacg cccggcacac ctgggtgacc
2881 tcgagcaggc cgagcaggc cggcgaaccc atgcgggcca gctctcgacg acgcagcg
2941 cgcgcggccgg cgatggcccg ctccgggtgc gggtagcggc tcgcccggc gatcagcgcc
3001 cgagccacgg cggcgggctc ggttcgagc agttcgggc cggcaggccc gtctcgatc
3061 tgctggatga ccccgccgc ggcaccaac agatccggca catacgccga ggtacccaaag
3121 acatgcatga gcccgtggc caccgcggc ttgtccgc gctggccag gtaccagtt
3181 tcggggcact gcccctact gagccgcgg taggcccacca gtcggccgc gggatgggg
3241 gcatacgaca tccagtcag cagcctggc acgaccccg actgcaccccg tccgcggccgg
3301 ccgcgttgc tgaccaacgc cgacatgtt tcaacgcgg tctgggtcc ctgttagcccc
3361 agcgcggcact gcccggccccc cggccctcc aacgtcatgc cgtggcgat ctccaaacccg
3421 gtcggccga tcgattccag cagcggttgc tagaagagtt tgggtgtaa ctgcacacc
3481 cgcacgttct gcttcttgc ttcccccgc agcaccccg cgcacatgtt tcggccatcg
3541 ggcggatgt gggccgcgc cggcagccag cgcactgcct cctcgatcc gggatgggg
3601 agcaggtggg tgcgttgc ggcgtcaac tgcagtccgt gctcgacccg cctgaggaaac
3661 tcatacgacg cggcatgtt cggcgttgc tcacgcccga tggccgc ttcgcccac
3721 gcccacatg cgtccaccgt ggacgcacc cgtaaacgact cgtcgatcc ggcacacc
3781 agctgcagta gctgtacggc gaactccacg tcgcgcacat cggccgttgc gagggttgc
3841 tcggccgcggc ggacatcgcc gggcaccagc tgcgttgc ggcacccgc ggcctgcacc
3901 tcgaccacaa agtccgc ctcgcaggc cggcaccacca tcggcatcaa ggcggcagg
3961 taacgctgc caagtccgc gtcgcacacg actggccgt cttcagcaac cgcctgaaac
4021 tcccaaggct tggcccgacg ctggtagtag gcgatgtgcg actcgacccgt acggaccagg
4081 tccccgttgc gcccctccgg acgcaggcgc ggcgttgc acggaccaggc

4141 accgcatca tctcgctggc cacgcgcgcg ttgcgcgggt cggagcgctc ggcaacgaat
4201 atgacatcg a cgtcgctgac gtatgtcgt tcgcgcac cgcacttgcc catcgatg
4261 accgcccaggc gcgggtggcgg gtgcgcggc cacacgcgtc cctggccac gcgcagcgcc
4321 gcccgcaggc cggcgccggc ggcgtccggc aggcggtggc ccaccacggt gaatggcagc
5 4381 accgggtcg cctcgaccgt cgcggccagg tcgagagcgg ccagcattag cacgtatcg
4441 cggfactggg ttgcataatcg gtgcacgagc gagcccgca taccctccga ttccctcgacg
4501 cactcgacga acgaccgcgt cagctggta tgggacggca gtgtgacctt gcccgcagc
4561 aatttccagg actgcggatg ggcgaccagg tgcgcggcc acgcgcgcg cggccgcgc
10 4621 accgagaaca gccgcgcgcg cagactcgat tcgcgcagca gagccgcgtt gagctcgcc
4681 catccggtgt ctggattctc cgcacgcggc atcaaggcgc gcgcgcggc atccgcgtcc
4741 ggagcgcgtg acagcgcacca cagcaggctg acgtgcgcct gatccctgtg ccgcattccac
4801 cccagctgag ccagacgcgc accagcaggg gggtaacta atccgagccg gccaacgcgt
4861 gccaactcg gcccgcgt ggcgagttt gtcacgcac cgcggtagc gcaaagcgc
15 4921 tcggcgtcgatcaaccggta agatctggc tacagcgcaca ggttagtgcg cagctcgat
4981 ggcgtgacgt ggctcggtt gttcgccac tccgtgcgt tgcgcgcaaa gaaaaagtca
5041 aaaaacgtgtcccccaggc ctccgcgcgc agttcgagg cctccatggc gcgcgcgc
5101 ctatccaaac tggacggcaa ttctcggtac cccatcgctc ggcgttcctc ggggttgagg
5161 tccctatacg tgcctcggtc ctgcgggccc agcactgtac cttctctac accccgcaat
20 5221 cccgcggcca gcagcacggc gaatgtcaga tagggattgc acggcgaatc agggctgcgt
5281 acttcgaccc gcccgcacga ggtttgtgc ggcgtgtaca tcggcaccgc cactaggcgc
5341 gatcggttgg cggcccccgc cgcgcgcgc gttggcgctt cgcgcgcgc caccagccgc
5401 ttgttaagatgtgacccactg atttgtgacc ggcgtgtatc cgcaaggctg tcggcaggatc
5461 cccgcgtatc acgatttacc cacttcgcac agctgcacgc gatcatcgc gctgtggaaac
25 5521 ggcgtgacat caccctcgaa caggctcatg tgggtgtcga tcggcgcgc cgggtgcgt
5581 cccatggct tgggtgtcga cgcgcgcgc ggcgcgcgc ccagcgcgc tcgttgcgt
5641 acgtacggcaggatgtcgtatc gttgtcgcgc atcgcacagag cgtcgccaa cccgcaggctg
5701 atccctgtc ggcgggtgc gcctcgatc tggctgtactt ccaccggatc gcccgtatc
5761 tccaggccat cgcgcgtg cggcgcgaaag tcaaggcgg agtgcgcac cgcgttgcgt
30 5821 aaatagccgg cgttgcac cgggacggc accgacccgt cctcggtcc gggcttgagg
5881 aggaagaact cgtttcggtt atgcacgtatc caggagaagc cgcgttgcgc ggccttcgt
5941 acgtgcgcgc gcaacacgtt cgcgggtgc gcccacgcac ggcgcgcgc cggcatggtg
6001 atgtcgcaaa acatccgcgc tgatgtggg tggccggaaac tggggccca gggcagcacc
6061 tggaaaggatc acgggtccgg gtgcgcaccatc gtatggatt cgcgcgcgc cggccaaaggccc
35 6121 tcgatcgagg atccgtcgaa cccgtatc tccgtcgatc cgcgcgcgc tccggctggg
6181 gcgatggcga cccgtatc gaaaccgcgc acgtcgatc accacagccg gacgcaggcgg

6241 atgtcgcggtt cttccagggtt acgaagaacg aattcccttct gtccgtccat acctcgaaaca
6301 gtatgcactg tctgttaaaa ccgtgttacc gatgcccggc cagaagcggtt gggggggggc
6361 ccgcgaagggtt agtgcgcgggtt gagttcagggtt cgcgcaccgc agactcgtcg gggcaagggtt
5 6421 cccgtcgaga aaatagtgcata tcaccgcaga gtccacacac tggttgccat cgaacaccgc
6481 agtgtgttgg gtgccgtcga aggtgatcag cgggtgcgc accgtggcggtt ccaggctac
6541 cccggactga tacggagtgg ccgggtcggtt ggtggtggac accacgacga ccttgcacgc
6601 cccggccggc gcccgggggtt gcccgggtt cgcgcgtcga cgttgcgc accggccaca ggcgcacag
6661 atcgcgaaaa gccggatccgg tgaactgccc gtagctaagg aacggggcga cctgacggat
10 6721 ccgttggtcg gcccgcaccc aggccgtgg atcgccgggtt gtggcgcat cgcacgcaccc
6781 gaccgcgttg aacgcgtcctt ggtcggttgcg ttagtgcctt tctgcatttcc ggcgcgtcata
6841 gtcgtcggca agcaccagca agtcgcggc gtcgtcgcg cgcgtcagcc ccaagcagacc
6901 actggtcagg tactcccaggc gctgagggtt gtagcgcgc ttgtatggc ccgtcgatgc
6961 gtcggcgttag ctccggccac gtcggatccga cgttcttcc ggcgttgcgca ccaggggtt
15 7021 aaccaggggcg tggtagcggtt tgacccactg ggccgagtcg gtggccagag ggcaggccgg
7081 cgagcggggcg cagtcggccgg cgttagtcattt gaaagcggtt tgaaatcccg ccattggct
7141 gatgtttcc tcgattgggc taacggctgg atcgatagcg cgcgtcagga ccatcgcccg
7201 cacatgagta ccgaaccgtt ccaggtaaac ggtgcccaac tcgggtccgtt agctgtatcc
7261 gaggttagttt atctgtatcg cacctaaccgc ttggcgaacc atgtccatgtt cccgtcgcac
20 7321 ggacgcggta ccgcattttgg ccaagaagctt gaagcccatc cggtaacac acgtctggc
7381 caactgcggg tagacctgtt cgcgtgggtt gacaccggcc ggactgttagt cggccatcgg
7441 atcgccggggtaacgcgtcga actcggcgtc ggtgcgcacac cgcacgcag gggtcgagtg
7501 gcccacccctt ctcgggtcga agccaccag gtcgaagtgg cggagaatgtt cgggtcgcc
7561 gatcgccgggtt gccatagcgg cgcattgtc gaccgtcgc gcccgggtt cccaggatt
25 7621 gaccaggactt gctccgaatc gctgtccgtt cgcggggacgc cggatcaccgc ccaacttcgc
7681 ttgtgtccca cccgggttgggtt cgttagtcgcac gggacggac accgtcgccgc agcgtcgc
7741 gcaatttcg ctgggtcggtt cgcgttgcgc acgttgcgc acgttgcgc
7801 cacgaccggc gcaccggggg ttggccggc gcccgggtt tcagtcgcgc cggccaaacgg
7861 gggcgctgtt agggcgactc cgcgcggcagca aacccgaag gacagcagcgc cgcagactcaa
30 7921 cggctcgccg cgcacatgg cgcgcattgtt ctcaccggcg aatacctgtt acggcgca
7981 atgatcacac cttcggttct tcgccccgtt agcacttggc gccgcggc ggcgtgggt
8041 cggcgattaa atacgcgcgtc acgtactgtt caatgcgcgtt gtcgcctgg aataccacc
8101 tggctgggtt cccgtcgaaat gtcagcaacg aaccgcgaag ctgggtcgcc aggtcgaccc
8161 cggccgttgcg cggcgtcgcc gggcatggg tgggtggatac caccaccgtc ggcacttaggc
35 8221 cggcgccga gacggcatgg ggctgacttg tgggtggcac cggccagaac ggcgcagggt
8281 ccagcggcgc atcaccgggtt aacttccgtt agctcatgaa cgggtcgatc tcccgccgc

8341 ggccgttcc gtcgatgacc ttgtcgcgat cggtaaccgg gggctgatcg acgcaattga
8401 tcgccacccg cggtcacccg gaattgttgt agcggccgtg cgagttccgaa cgcacatgtaca
8461 tgtcgccag agccagcagg gtgtctccgc gattgtcgac cagctccgac agcccgtcgg
5 8521 tcaagtgttcc cacagattt ggtgagttaca ggcgcataat ggtgcccacg atggcgtcgc
8581 tataactcag cccgcgcgga tccttcgtgc ggcgcggcct gctgatcctc gggtgtcgg
8641 ggtcgaccaa cggatcgacc aggctgtggt agacccgtcgc ggcttggcc gggcggcgc
8701 ccagcgggca gcccgcgttcc ttggcgcagt cggcggcata ttgttgaac gcgtcctgg
8761 agcccttggc ctggcgcagc tccgcctcga tgggatcggc attggggatcg acggcaccgt
10 8821 cgagaatcat tgcccgcacc cgctgcggaa attcctcggc atacgcggag ccgatccggg
8881 tgccgtacga gtagccagg taggtcagct tgcgtcgc cAACGCCGCG cgaatggcat
8941 ccaggtccctt ggcgacgttcc accgtcccgaa catggggcag aaagtcttgc cccatcttgc
9001 ccacacagcg accgacgaat tgcttggctcgat gtgcgcaca ccctccggc
9061 tgttagtcaac ctggcgcgttcc gcccgcagcc ggtcggttgc ggcacatggag ttgcaccaga
15 9121 tcgcccggccg ggacgacgccc accccgcggg ggtcgaaacc aaccaggatcg aacccttcgt
9181 gcacccgttcc cggcaatgtc tggaaagacgc ccaaggcggc ctcgataccg gattgcgg
9241 gtccacccggg atttatgacc agcgaaccga tcttgcgtcc cgtgcggga aagcgaatca
9301 ggcgcagcgc cgcacgtca ccatcggggc ggtcgtagtc gaccggtaca gcgagcttgc
9361 cgcataacgc gcccggggg atctttactt gcggtttga cgaccggcac ggtgtccact
9421 ccacccggcttcc gcccgcgttcc ggctccgcacc tacgagcgcc tcccccgcacc acgcggatgc
20 9481 agccccacaag aaccaacgc accggcggcga ggcgcggccca gatcaacagc atgcgcgcga
9541 tcttgcgtcc gcgagacagg ctcatgccc caatgtcgcc agagcagacc cgagatccgt
9601 gccagcggcc accgtcgccc gactaaccgg ccgtcgccag cagtcctgc accgcgcgt
9661 gcgaaactcgt cggccatccc ccatacgatcc ggtaaacagat ccggcaaga caccgaccgg
25 9721 tcgaccggat ccggcacggg cgcgtcgcc tcggcggtgc acaactgcga catcaggatgg
9781 ggcgtggcac cccgtccacg ccggcatggt gcacccgttgc catcgcccgaa gggcgatccc
9841 cgtatgcgtcc accccttcg acgaaccat ctccacggc ggtcgccggc agcgacgcga
9901 tgtggccgca gatctccgag agttcgccc gcccggccgg cgaacggcaac ccgtgcgt
9961 gcaagtgcacg atcgatgtga ggttcaaggt tcagcgact gtcggcaacg ttttccgaa
30 10021 accgcggcct cgccttgatc tggagtca gacgcgtcaccg cagccggtaa aaggcgtaac
10081 ccatgctcgatc gcaaaacatgc atgggctgatc tggacgttgc cagacacagc aactggcgatc
10141 caggccactg agccgcgtca tgcgcgtatgg tgcgcgtatgg gggcccccgg ggcgcgtctga
10201 ggggaagaag tggcagactg tcagggtccg acgaaccgg ggaccctaacc gggccacgag
10261 gatcgaccccg accaccatgg gggacagtga tgtctgagca gactatctat gggccaata
35 10321 ccccccggagg ctccggcccg cggaccaaga tccgcacccca ccacctacag agatggaaagg
10381 ccgacggcca caagtggcc atgctgacgg cctacgacta ttgcacggcc cggatctcg

10441 acgaggccgg catcccggtg ctgctggtcg gtgattccgc ggccaacgac gtgtacggct
10501 acgacaccac cgtgccgatc tccatcgacg agctgatccc gctggccgt ggcgtggtgc
10561 ggggtgcccc gcacgcactg gtcgtcccg acctgccgtt cggcagctac gagggggggc
5 10621 ccaccggccgc gttggccgccc gccacccgggt tcctcaagga cggcggcgca catcggtca
10681 agctcgaggg cggtgagcgg gtggccgagc aaatcgccgt tctgaccgcg ggggcattcc
10741 cggtgatggc acacatggc ttaccccgca aagcgtcaa cacctggcggc ggcttccggg
10801 tgcagggccg cggcgacgcc gccgaacaaa ccatcgccga cggatcgcc gtcgccgaag
10861 cgggagcggt tgccgtcgatggatggatggatggatggatggatggatggatggatggatggatgg
10921 gcaagcttac cattccgacg gtcgggatcg ggcgtggccca actgcgac gcccagggtcc
10981 tggtatggca ggacatggcc gggtcagcg ggcgaagac cgcccgcttc gtcaaacgggt
11041 atgccgatgt cggtggtgaa ctacgcccgtg ctgcaatgca ataccccaa gaggtggccg
11101 gccccgttatt ccccgctgac gaacacagtt tctgaccaag cggatcgac cggatcgccg
11161 ggcattgcgg tggcccccgtg gatgccgtcg acgcccggatt gccggcgccgg acgcccggc
11221 gggacccatc ggcgtcgctg tcgcccgtt agccccgggt gagcccagac attcgatgt
11281 cccaacacca tccgcccacag cccaaatgtat gtcggactct atgcgtgcct atcccgacc
11341 aaccaccacc gcgccgacgc atcatgaccg gaggcgaaga tgccagtaga ggcgcccaga
11401 ccagcgcgcc atctggaggt cgagcgaag ttgcacgtga tcgagtcgac ggtgtcgccg
11461 tcgttcgagg gcatgcccgc ggtgggtcg tcgagcgtcg cggcaccac gcaagtcgac
20 11521 gccccgttact tcgacacacc gtcgcacgac ctggccgcaccagatcac ctggccgc
11581 cgcacccggcg ggcggcgcgc cggctggcat ctgaaatgtc cggccggacc cacaaggcgc
11641 accgagatgc gaggcaccgt gtcggcatca ggcgacgtcg tgccggccga gttgtggat
11701 gtgggtctgg cgatcgccg cgaccagccg gttcagccgg tcgcccggat cagcactcac
11761 cgcgaaagcc agatctgtat cggccgggg ggcgacgcgc tggccggatt ctgcaacgc
11821 gacgtaccgc catggccggc cggggcattc cacggccgtg gtgcagcggca aacggccct
25 11881 ggcgaacagc agtggcgca atgggaactg gaactggatca ccacggatgg gaccggcgat
11941 accaaggtac tggaccggct agccaaacccg ctgtcgatg cggccggccgc acctggccgc
12001 cacggctcca aactggcgccg ggtgctcggt ggcacccctc cgggtgagct gccaacggc
12061 cgcagccgc cggccggatcc agtacaccgc ggggtgtccg agcaagtcga gcagctgtcg
12121 ctgtggatc gggccgtcgcc ggccgacgcc tatgacccgc tgccaggat ggcagatgc
30 12181 acccgcaaga tccgcagctt gctgacggat tcccgaggat cgtttggcct gaaggaaagt
12241 gctgtggatca tcgatgaact gctgagctg gccgtatgc tggccgtac cggggacgc
12301 gaggtactcg tgaccgcata ccagcgcgaa ctggacgcgc tggccggat gctggatgc
12361 ggcggggatc gcgagcgcct ggttagacggg ggcggccggc gataccagac cgggtgtccg
12421 cgtactga tcgcattgcg gtcgcagccg tacttccgtc tgctcgacgc tctagacgc
35 12481 ctgtgtccg aacgcgcata tgccacttct gggggatcgccaccggat aaccatcgat

12541 gcggcctacc ggcgagtccg caaagccgca aaagccgcaa agaccgccgg cgaccaggcg
12601 ggcgaccacc acccgacga ggcattcac ctgatccgca agcgccgaa gcgattacgc
12661 tacaccgcgg cggctactgg ggcggacaat gtgtcacaag aagccaaggt catccagacg
5 12721 ttgtctggcg atcatcaaga cagcgtggc agccggaaac atctgatcca gcaggccata
12781 gccgcgaaca ccgcggcga ggacaccctc acctacggc tgctctacca acaggaagcc
12841 gacttggccg agcgctgccc ggagcagctt gaagccgcgc tgcgcactt cgcacaaggcg
12901 gtcgcgaaag cacgggattt agccgcggcag gggccggacga gttggctgt aagccggatt
12961 ctgtccgcg ccgcacacgc caagctaacg gcccacggc ggcgaccatc catctggaca
10 13021 caccgttacc ggggtcctcg agcggctac ccgcaggc gggcgagcaa ccctcaagcg
13081 cctgcgcggc cgcaacttcg gtgcggccctt ctggccttgc ctgcgggtgg gggttgccta
13141 gccaccccgga tcacccggaa tgctggcgc ctcttaccgc accgttac ccttgccacc
13201 acgaggatgg cggctctgttt tctgtggcac ttcccgca gtcacccgtt attggcgta
13261 gcaatcaccc tgcgtgtga agtccggact ttctcgact cgacgtgaa ctcgtgaat
15 13321 ccacacaaggc cctacgcgag ccgcggccgc ccagccaaact catccgcgac gaccacgcta
13381 ccccgctggg cgggtgcgcg gccagttgtga ccgcgtggacg acacggctag tcggacagcc
13441 gatccggcgg gcaatccatca tcgtggactg gtacacgggtt gggacaaacg cgtcgactcc
13501 ggccactggg acgccatcgc tgccgaggc agcgagttacg gtggcgcact gctacccgtt
13561 ctgatcaccc cggcgaggc cggccggctg cgcaagctgt acgcccacga cggccgttt
20 13621 cgctcgacgg tcgatatggc atccaagcgg tacggcgcg ggcagtatcg atattccat
13681 gccccatc ccgagtgtac gagcgtctca agcaggcgctt gtatccaaa ctgctgccga
13741 tagcgcgaa ctgggtggcc aaactgggccc gggaggcgcc ctggccagac agccgtatcg
13801 actgggtggc gagctgtcat gcccggccaa acaccgtatc cacagcgctt atgttgaagt
13861 acggcaccaa cgactggaaac gcccataacc aggtatctta cggcgagttt gtgttccgc
25 13921 tgccgttgc gatcaacccg agcgatccgg aaaccgacta caccggccgc gagttccctgc
13981 ttgtcgaaca gggcctcgc gccaatccc ggggtaccgc aatgcaactt ccgcaggggac
14041 atgggtatgt ttccacgacc cgtgtatccgc cgggtggac tagccgttgc tggcggcat
14101 ctccagtgcg ccatggcctt tcgactatccgc tttccggcga acgctatgcc atggggcttga
14161 tcttcacga cgcagcctga ttgcacgcata tctatagata gcctgttgc ttcaccaatc
30 14221 gcaccgacga tgcccatcg gcgtagaact cggcgatgtc cagcgatgcc agatcaagat
14281 gcaaccgata taggacgccc gaccggcat ccaacggcag ccgcacaaac atttgtatcg
14341 gctgtacatg tgacaccacc agcaccgtcg cgcctcgta gccaacgtatc atccgtatcac
14401 gtcggccgc aaccggccgc agcacgttgc cgaagtttccacccggg ggcgtgtatgc
14461 tgggttcctg cagccagcga cgggtgcgtt cgggtgcgc ttctggcc tccgcgaacg
35 14521 tcaagccctc ccaggcgccg aagtccgttgc gacccaggcgtt cgcacgttca accacgttca
14581 gggccaggcgc tctggcggcgtt caccggcgg tgcgttgc cgcgtgtac ggcgaggaga

14641 ccaccgcagc gatcccgccg cgccgcgcca gataccggc cgccgcacca acctggcgcc
 14701 accccaccc gttcaacccc gggttgccgc gccccgaata gcggcggtgc tccgacagct
 14761 cctgtcgccc gtggcgcaac aaaagtagtc gggtgggtgt accgcgggacg cccgtccagc
 5 14821 cgggagatgt cggtgactcg gtgcacacga ttttggcagg atccgcaccc gccgcagccg
 14881 attgcgcggc ggcgtccatc ggcgttgc ccaaccggc tgcatacgtg ttccgggcac
 14941 gggaaaccca ctcgttagtt atcctgcgaa actgggacgc caacgcctga gcctggacat
 15001 agagcttcag cagatccggg tgcttgcacct tccaccggcc ggacatctgc tccaccacca
 15061 gcttggagtc catcagcacc gcggcctcgg tggcacctag tttcacggcg tcgtccaaac
 10 15121 cggctatcag gcccggatcg tggcgacgt tggtcgctgc cggccgcatac gcctgcgttg
 15181 actcggccag cacggtgag tgatcggcgg tccacaccac cgcggcgat cggccggcgc
 15241 cgggattgcc cgcgcgtcc cgcgtggctt cgtacgacaac tttactctt caaatccctt
 15301 gagccgaac aagatcgctc cgcattccgg gcagcgcacc acttcaccc cggccggccgc
 15361 ctagatctgg cccagctcgc cgcggccat ctcgatccgg caggcaccac atcgatgacc
 15 15421 ttgcacccgc cggcccccgt gcccgcctcc gcccgcgtgt ctgcgttgc gcccgcgaag
 15481 ctggggatca agtgcgcgc tcagcatgtc ggcgttgcgt gaatgtggt gcccggcttg
 15541 gtcgatttcg gcaagtgcct cgtccaaagc ctgcgtggcg gcccgcagg cggccggca
 15601 cgcgtggagc gcccgcact cggcggtctg ttgacgcgtc agtcctcgc ggcgttccag
 15661 cacccgcacg agggcatctt cccaaactggc ttgacggcgt tgcaagctgt cgcgttgc
 20 15721 ctgcagatca gccaattgtct tggcgccgt tgacccgaa gtgagcaacg accggcccg
 15781 gtcgcacgc ttacgcaccc catcgatctc cgcgttgc ggcgttgc ggcgtccaa
 15841 gtcctccgc cgcatttcgca gggccgcatt cctgtcggt gcccgggtgt gtcggccctg
 15901 cacctgttgg taagccgcgc gctgcggcag atgggtagcc cgcgtgcga tccgggtcag
 15961 ctcagcatcc agttcgcca attccagtag cgcgttgc tgcgttgc cggccatc cggccatcc
 25 16021 gctgtatctc tcccaatgtc ttgcgttgc gtcgttgc ggcgttgc ggcgttgc
 16081 cgcacccggca ggcgcgcgc gaaatgagac cgcacaccc cggccggctg gcccgcaccac
 16141 gggaaatttcg tcgcgttgc cgcgttgc atcaggccca ctgcgttgc tcggcaatgc
 16201 tcgtcggtgt gatgtgtcg cagatccggc gtaacgtacg ctgcgttgc cgcggccggcc
 16261 acgggtggcaa gcaacggatc cccggccgc cgcgttgc cgcacccgcga caccaggcagg
 30 16321 tcgggatccc cggccggccg cacaccggc gcagtcggcg gcaacggcc ctccagacgg
 16381 gcaacaaagg tgccgttgc ggcgttgc caatccggcc taaccggctg
 16441 cgcacccggcg gtcgttgc cgcgttgc cgcgttgc gtcgttgc
 16501 ggcgcgttgc cgcacccgc ctcgttgc gtcgttgc gtcgttgc
 16561 tcctcggttgc cccgttgc cgcgttgc cgcgttgc
 35 16621 gccaggaact gcccggtacc cgcgcacactc cagctgcagt ggcgttgc gccgtatgg
 16681 cggccacccgg cctcaaagac cgcgttgc cccgttgc agtgcgttgc cggccatcc

16741 atgaccact tgtcgagatc ggccgctccg ggcacccgggt cgagaacggc gtcgacggtc
16801 agaccaacag cgtgtgccag cgctcgac acacccggcg acgcccagtc ggcgttggtg
16861 tgcgcggtaa acaacgagcg accggctccgg atcaggcggt gcaccagcac accctttggc
5 16921 gtgtggccg cgaccgtatc gacccacgc agtaacaacg ggtggtcac caatagcagt
16981 ccggcctggg gaacctggtc caccaccgc ggcgtcggt ccaccgcaac ggtcaccgaa
17041 tccaccacgt cgtcggttc gccgcacacc agacccacccg aatcccacga ctggcaaggc
17101 cgcggcggtt aggcctggtc cagcacgtcg atgacatgg ccagccgcac actcatcgcc
17161 gtcctccacg cttggccac tcggcgatcg ccgcccacccg cacggccac tccggcgca
10 17221 ccggccccc caggtaccgc gcttccaggc cgaccaagggt gtacccgcgg cgcaccgcaa
17281 ttcccttgct ctgcaaatacg ttcgtaatc cgtcagcatc ggcgtatgtt aacagtacga
17341 aaggggccgc accatcgacc acctcgaccc ccaccgatct cagtccggcc accatctccg
17401 cgcgcagcgc cgtcaaccgc accgcattcg ctgcggcagc ggcgaccgccc cggggggcgc
17461 agcaaggcgc gatggccgtc agttgcaatg ttcccaacgg ccagtgcgtc cgtgcacgg
15 17521 tcaaccgagc cagcacgtt ggcgagccga ggcgcgttccacccgcaat cggccageg
17581 accacgtttt cgtcaagcta cggagcacca gcacatggg cagcgagtca tcggccaacg
17641 attgcggctc gccgggaacc caatcagcga acgcctcgac gaccaccagg atgcgtcccg
17701 gccggcgtaa ctgcgacgc tgcgcggta ggtgcagcac cgagggtggg ttggcgat
17761 taccacgcac gacaagggtcg gcgtcgatcg gcacgtgcgc ggtgtccagc acgaacggcg
20 17821 gctttaggac aacatggtgc gccgtgtatc cggcagcgct caaggctatg cccggctcg
17881 tgaacgcggg cacgacgatt gctggccca cgggacttag gttgtgcagc aatgcgaatc
17941 cctccggccgc cccgacgcgc gggagcactt cgtcacgggt tctgcatgc cgttcagcga
18001 ccgcgtcttg cggccggatcg acatcgatcg tgctcgatgc gcccggccagc tccggcagca
18061 ggcgcggcgag ctgcggacc aaccattccg gggccgggtc atggcgacgc ttgacggcga
25 18121 agtccagcgcac gccggcgcg acatcgatcg caccgtggta gcgcgcgcg gcaagcgggc
18181 tagtgtctag actcgccaca ggcgtaaaca gtgtggcc ggtgtgcggg ccaagaatcc
18241 agagcaccgc cgcgcgttg tctacgcggc gacaaccgcg acatcacagg cagctaacag
18301 ggcgtcgccg ggtatgtatcg tcaggccaaag cagctgtgc tggtcgatgc gcacacggtc
18361 gaatggatgt cgtatggatcg cggaaagtc tgcggcgatcg agtgtgtgc tggtcaactg
30 18421 acagcggcga cgtgcgcag cggcgattc gatgggcac gtaagaagcc gatggctcg
18481 gcccggggag ctggccgagg cggtagttga tcgcgtatcg ccaggactg gcccggcaca
18541 agagaatgtt gttgcggacg tcctgaacaa tcggccgtt ttcgttgcacg gcatccgcag
18601 ccaaacgtgg gtgtcgatgc ggtgcgtt caccgggtaa agcgttcagc cacgtcgatcg
18661 gacaacggag cgtccaaatc gtccggcgcg cggatcacgc catggatcgatgc gcttaaccgc
35 18721 cgagtctat gaggatgcag cggcacaagc tttgcgtaccg gtcgcgcg gccggcaatc
18781 tcaacctctg cccggcgtag acgagccgca gcagctcgatcg caggcgatcg ttgcctcg

20941 ccggacaaggc caatgacggt cgagcagaaa ctgcgtgacc gctacggccg ccggcggtaa
21001 accaacaatca cggccaatac cgcagcccc gcctggacca cccgcgacag caccacggcg
21061 cggcgagat cggccaccc gggcgaccgg cggtgcggca aggtgggccc gatctgcaac
5 21121 tcatggtggt accgggtggg cccacccaggc cgcacgtcaa gcggcccgac aaacgcccgc
21181 tcgacgacac cggcggtggg gctggatgg cggcgccgt cgccgcggca ggcccggtacc
21241 gcaccgcggg gcgacccacc gaccacccggc gcgcagatca ccaccagcac cgccgtcgcc
21301 cgtgcgccaa catagttggc ccagtcatcc aatctgtct cagcccaacc gaatcgagaa
21361 taacgcggcg agcggtagcc gatcatcgag tccaggggtgt tcatggcagc atatcccagc
10 21421 accgcaggca cggcgctcga agccgcccac agcagcgca ccacctggc gtcggcggtg
21481 tttcgccca ccgactccag cggcgacgc gtcaggcccg ggccgcccag ctgggccccgg
21541 tcaacccgc acagcgacgg cagcaggcg cgcggccct cgacatcgac ggcgtccaac
21601 aggtccgata tctggcgccg ggtgcgcgc agcgaagtgc cgccagcgc tgccaggtg
21661 gccgtcgccgg tggccgaccc gggccaggac ctgcccggta gccgctgcag tgccgcccgg
15 21721 agcaagccca cggcgccgac cagcaggccg acgtgtaccg caccggcgac cggccgtca
21781 cggtaggtga tctgctccag cttggcgccg gcccggaccga acagggccac cggatgaccc
21841 cgtttgggtt cgccgaacac gacgtcgagc aggccggca tcagcagcgc gacggccctg
21901 gtctgccagg tcgatgcaaa cactccggca gcgtcgacca cgtggctac gctcagctat
21961 ttatgacctc atacggcagc tatccacgt gaagcgccca gctacccggg ttggccgaccc
20 22021 gttgaacccg gccgcaatgt ttttgcggc agcgaatgtc atcatgcagc tggcagtggcc
22081 ggggtcgccg ttttgcgtgc tggaaagccc ggtggacagc ggcaacgtct acaagcatcc
22141 gtcaagcgccg gcccggacca cggcaccta cttggcggtg gggaccatcg ggacggaaatc
22201 cgaccgagcg ctgatccggg gtggcggtt ctttttttttggccgaccc cggcagggttc ggtcgacggc
22261 ctcgagccca gtgtcctata acgccttcga cccgaagttt cagctgtggg tggcgccgtg
22321 tctgtaccgc tacitcggtt accagcacga gtttttttttgcggccggcggccactcg aagatgcccac
22381 cggcgacgccc gtcttaccaag acgcacccaaacg gtttagggacc acgtgcagg tgccggaggg
22441 gatgtggccg cggaccggg tcgcgttgcga cgactactgg aagcgctcgc ttatgggtt
22501 gcagatcgac ggcgggtgc gcgagcatct tcgggggtt gctcggttag ctttttttttcc
22561 gtggccgttgc cggcgccgtt ccggggccgtt caacctgttt ggcacgcgg gattttggc
30 22621 accggagttc cgcgcgtatgc tgcagctggatgtt gtttttttttgcggccggcggccggcc
22681 gtttttttttccgttgcac gtttttttttgcggccggccggccggccggccggccggccggcc
22741 ctttttttttccgttgcac gtttttttttgcggccggccggccggccggccggccggccggcc
22801 ctgatagagc cggcccgagt gtggccgttgcac gtttttttttgcggccggccggccggcc
22861 ggcagggttca cgcgttgcac gtttttttttgcggccggccggccggccggccggccggcc
22921 atcaacgtgc ttttttttttgcggccggccggccggccggccggccggccggccggccggcc
35 22981 gcatacaggt cggccgttgcac gtttttttttgcggccggccggccggccggccggccggccggcc

23041 gctgtctgtc gcgccgcgcgt gtcgtttga aaccaggc gtcacccac gaccacgaca
23101 ctgccccatc cggcgccccg ccgacaacga agcacagcta gccgggtggc gcggacggga
23161 tgaaccgccc gaccgctggt gtgtaaaacc agagctctac cgctgagcta cgcccccatt
5 23221 accggccgcag gctacacgccc ttgcggccaa gcacccaaaa ccttaggccc taagcgccgc
23281 cagagcgtcg gtccacagcc gctgatcgac aacttcaccc ggctgcttca tctcggcgaa
23341 ccgaatgatc cctgaccgat cgaccacaaa ggtgccccgg ttagcgatgc cggcctgctc
23401 gttgaagacg ccgtggccct gactgaccgc gccgtgtggc cagaagtccg acaacagcgg
23461 aaacgtgaat ccgctctgctc tcgcccagat cttgtgagtg ggtgggggc ccaccgaaat
10 23521 cgctagcgcg ggcgtgtcg ctgttcaaaa ctggggcagg ttagtacgcgca actgggtccag
23581 ctgcgcctgg cagatgcggc tgaacgcaca cggaaagaac accaacagca cggttcttgc
23641 accccggtag ccgcgcaggg tgacaagctg ctgatttctgg tcgcgcacg tgaagtctagg
23701 ggcgggtggct ccgacgttca gcatcagcgc ttgccagccc gcgatttcgg ctgttaccaat
23761 ctgcgtggcgc tccagttgcc cagattgacc gacgaggctg gcatcagccc agctgtggc
15 23821 gccgcctcgaaatctcgcc gggcaataca tggccgggct ggccggctt gggcgtcacc
23881 acccaaatac caccgtccctc ggcgagcggg ccgatcgcat ccacccagggt gtccaccaaa
23941 tcgcgtcgcc catacgcaca ccacaacagg acgacatcg tgcacccgtc ggtgtttca
24001 tcgagcaact ctcccccgcgc cgcitctcg atggccgcgc ggatgtcg tcgggtgtct
24061 tcgtcccagc cccattctcg gataagtttg ttcgttgaa tgcccaattt gggggcgttag
20 24121 ttgcaggcgat gatccgcgc gaccaccgtg gaaaccttccctt cagtcgcgc gggccatgtg
24181 cacaccgtcg ccatggccat tatcgtcgca cagccagaac cggccaccc gcccgcctca
24241 gaaggcggcc acgcacatcg tcaatgcctt tgcgttggc gatcaacccg
24301 ccgggtgaat tccgcgtcg acgcgtgcgc accgatggca ttgccaccc ggcggccgc
24361 gtcgacatat gcggtgagcg catccccag ttgcgcggac agcgcggcgc tcagactgcc
24421 ttagaccgtc gaggcactgt tgcgttggc gtcgttggc ggaccccttccgg tggccgggt
25 24481 gttgcggccccc tgattgaacg cggccacgtt ggcgttccacc ttgcgtatgg cgttccctgt
24541 ggtggccgc acgcgtcac acgagggtcg aatgccttgc gtcgtcagcg atgttggcg
24601 ctgcgactcc cggatgtcg acgtcgccgc cgaaggccgac accgacgcgg acaccgacga
24661 gcggtaggcc ggtgcgacgt tgggtcgccatggccatggc ttcgtcgatggc cagttggata
24721 tccgacgatc cccatcagca gcagcgcgat gcagccgagc gccaggccgc ctgcgcctggg
30 24781 gagctcccccc ccgtgcctcg gaggcacggc ggcgcacccatccg atgagcacgg catgtgagg
24841 tacctggtcg cagcgcgacc ggcgtggccg tgggtgtcg cgcacccatccg gatccgcgc
24901 gagtgccgcgt atccgcgcgc gacgcgggtg cggcacgata gggggacgac catctaaaca
24961 gcacgcacgc ggaaggccgc cacctacagg agtagtgcgt tgaccaccga ttgcgcgc
35 25021 cacgatctgg cccaaaactc aaacagcgcacgacgaccaccg accgagttcg ggtgtatccgc
25081 gaggggtgtgg cgtcgatgtt gcccgcacatt gatccgcggag agacccatccgatggatggag

25141 tcctttgaca cgctgctgca acgctgcggc ccgtcgccgg cccgtaccc gatgtgcgg
25201 ctgctagagc gggccggcga gcagcgggtg gccatccgg cattgacgtc taccgactat
25261 gtcaacacca tcccgaccga gctggagccg tgggtcccg ggcacgaaga cgtcgaacgt
5 25321 cgttatcgag cgtggatcgat atggaatcg gccatcatgg tgcaccgtgc gcaacgaccg
25381 ggtgtggcgc tgggtggcca tatctcgacc tacgcgtcg ccgcggcgct ctatgaggc
25441 gggttcaacc acitcttccg cggcaagtcg caccggcgcg gggcgatca ggtgttcatc
25501 cagggccacg cttccggg aatctacgcg cgccgcctcc tcgaaggcgc gttgaccgc
25561 gagcaactcg acggattccg ccaggaacac agccatgtcg gcggcgggtt gccgtccat
10 25621 ccgcacccgc ggctcatgcc cgacttctgg gaattccca ccgtgtcgat gggttggc
25681 ccgctcaacg ccatctacca ggcacggtc aaccactatc tgcatgaccg cggatcaaa
25741 gacacctccg atcaacacgt gtgggtttt ttgggcgacg gcgagatgga cgaacccgag
25801 agccgtggc tggcccacgt cggcgcgtg gaaggcttgg acaactgac cttcgatc
25861 aactgcaatc tgcagcgact cgacggcccg gtgcgcggca acggcaagat catccaggag
15 25921 ctggagtctgt tctccgcgg tgccggctgg aacgtcatca aggtgttgc gggccgcgaa
25981 tgggatgccc tgctgcacgc cgaccgcac ggtgcgcgtt tgaatttaat gaatacaaca
26041 cccgatggcg attaccagac ctataaggcc aacgacggcg gctacgtgcg tgaccactc
26101 ttcggccgcg acccacgcac caaggcgtg gtggagaaca tgagcgacca ggatatctgg
26161 aacctaacc gggcggcca cgattaccgc aaggtttacg ccgcctaccg cgccgcgc
20 26221 gaccacaagg gacagccgac ggtgatctg gccaagacca tcaaggcata cgccgtggc
26281 aagcatttcg aaggacgcaa tgccacccac cagataaaaa aactgaccct ggaagacctt
26341 aaggagttc gtgacacgcg cggattccg gtcagcgcacg cccagcttga agagaatccg
26401 tacctgccgc cctactacca cccggcctc aacgccccgg agattgtta catgtcgac
26461 cggcgccggg ccctcgaaaaa ctttgtccg gagcgcagga ccaagtccaa agcgctgacc
26521 ctggccggc gcgacatcta cgccgcgtg aaaaagggttgc ctgggcacca ggaggtggcc
25 26581 accaccatgg cgacgggtcg cacgtcaaa gaagtgttgc ggcacaaagca gatcggcccg
26641 cggatagtcg cgtatcccg cgacgaggcc cgccaccccg ggtggactc ctgggtcccg
26701 tcgctaaaga tctataaccg caatggccag ctgtatccg cgggtgacgc cgaccatgt
26761 ctggcctaca aggagagcga agtcggcag atcctgcacg agggcatcaa cgaagccggg
26821 tcgggtggc cgttcatgc ggccggcacc tcgtatgcgac cgccacaacga accgatgatc
30 26881 cccatttaca tcttctactc gatgttgcg tccagcgcg cggcgatag ctctggcc
26941 cggccgacc agatggctcg aggggttcgtg ctggggcaca cccggccggc caccaccctg
27001 accgggtgagg gcctgcaaca cgccgacggt cactcgatgc tgctggccgc caccaccccg
27061 cgggtgggttgc cttacgaccc ggcctcgcc tacgaaatcg cttacatcgatc gaaagcgg
35 27121 ctggccagga tgtgcgggaa gaacccggag aacatcttct tctacatcac cgtctacaac
27181 gagccgtacg tgccatgcg cggccggc gggccggatc aacttcgatc ccggggcgatc gtcgtgggtt

27241 atctaccgct atcacgcggc caccgagcaa cgcaccaaca aggcgcagat cctggcctcc
27301 gggtagcga tgcccggc gctgcggca gcacagatgc tggccgcccga gtggatgtc
27361 gccgcccacg tgtggcggt gaccagttgg ggcgagctaa accgcgacgg ggtggccatc
5 27421 gagaccgaga agctccgcca ccccgatcg cgccggggcg tgccctacgt gacgagagcg
27481 ctggagaatg ctggggccc ggtgatcg cg gtgcggact ggtgcgcgc ggtccccgag
27541 cagatccgac cgtgggtgcc gggcacatac ctcacgttgg gcacccgacgg gttcggctt
27601 tccgacactc ggccgcgc tcgcccgtac ttcaacaccc acgcgaatc ccagggtgtc
27661 gcggtttgg aggcggtggc gggcgacggc gagatcgacc catcggtgcc ggtcgccggc
10 27721 gcccggcagt accggatcga cgacgtggcg gctgcgccc agcagaccac ggatcccggt
27781 cccggggcct aacgcccggcg agccgaccgc ctggccga atctccaga aatctggcgt
27841 agcttttagg agtgaacgac aatcagttgg ctccagttgc ccgcggagg tcggcgtcg
27901 aactgttggc cactgtgcc gattcgctc tgccgggtt gaagcgtac tcggccggc
27961 tggccaccga ggcagttcg gccatgcaag aacgggtgcc gttctcgcc gacctagaag
15 28021 cgtcccagcg cgccagcgtg ggcgtgggg tgccagacggc cggtcaac ttgtcgaaat
28081 ggtgcacga cccgcacagt gacgtgggtt ataccgcga ggcattcgag ctggccccc
28141 aggtctgac gcgcacggatc ggcgtgcgc agaccgtgga catggtgcgg gtcaccatgg
28201 agttctcga agaagtctgt cccctgtcg cccgttcga agagcagttt accgcctca
28261 cggtgggcat ttgaaatac agccgcgacc tggcattcac cgccgcacg gcctacggc
28321 atgcggccga ggcacgagggc acctgggaca gcccggatggc ggccagcgtg gtggacgcgg
20 28381 tggtaacgcgg cgacaccggc cccgagctgc tgcggccggc ggccgcgtg aattgggaca
28441 ccaccgcgc ggcgaccgta ctggggaa ctccggccgc cggccaaat ggctccaaaca
28501 ggcacggcga cagcgagggc ggcagccagg atgtccgcga caccgggtc cgccacggcc
28561 ggcgtgcgt gaccgacgtg cacggcacct ggctggggc gatcgctcc ggccagctgt
25 28621 cgccaaaccga gaagttcctc aaagacctgc tggcagcatt cgccgacgc cccgtggta
28681 tcggcccccac ggcgcacatg ctgaccgcgg cgccacgcg cgcgtgcgt ggcgttcgc
28741 ggtgaacgc cgtcgccggc tggcgccggc cgccgcggcc cgtgtggct agggacttt
28801 tggccgaacg cgcctgtat ggcgacgc cggccatgtt ggccctgcac accgacgtga
28861 tggcccccct agccgatgcc ggaccgacgc tcatcgagac gctagacgca tatctggatt
28921 gtggcgccgc gatgtaaatgttgcagaa agttgtcg tcatccaaac acagtgcgggt
30 28981 accggctcaa ggcggatcacc gacttcacccg ggcgcgtcc caccgcgc cgcgtgcct
29041 atgtccctcg ggtggccggc accgtgggtc aactcaacta tccgacgcgc cactgaagca
29101 tcgcacagcaa tggcggtgtca tagattccct cggccgtcgtt aggggggtcca gcagggggccc
29161 cggaaagata ccaggggcgc cgtcgacgg aaagtgtatcc agacaacagg tcgcgggacg
35 29221 atctaaaaaa catagcttac aggcccgtt tggggatatac aaaaaac ctaagacgag
29281 gttcataatc tggacaccgc cgccaaaccgc tcttcacagt gttcttttagt acacgtgtt

29341 gcgttgctcg cacccggaca gggttcgcaa accgagggaa tgggtcgcc gtggcttcag
29401 ctgccccggcg cagcggacca gatcgcggcg tggtcgaaag ccgcgtatct agatcttgc
29461 cggctgggc a caccgcctc gaccgaggag atcaccgaca cccgggtcgcc ccagccattt
5 29521 atcgtcgcccg cgactctgtt gcccaccag gaactggcgcc gccgatgcgt gctcgccggc
29581 aaggacgtca tcgtggccgg ccactccgtc ggcgaaatcg cggcctacgc aatcgccgg
29641 gtgatagccg ccgacgacgc cgtcgccgtc gccgcccaccc gggcgccga gatggccaag
29701 gcctgcgcca ccgagccgac cggcatgtct ggggtgtcg gggcgacga gaccgagggt
29761 ctgagtcgccc tcgagcagct cgacttggtc cggcaaaacc gcaacgcccgc cggccagatc
10 29821 gtcgtcgccg gcccgtgac cgcgttggag aagctcgccg aagacccgccc ggccaaggcg
29881 cgggtgcgtg cactgggtgt cgccggagcg ttccacaccc agttcatggc gcccgcactt
29941 gacggcttgc cggccggccgc ggcacacatc gcaaccgcgc accccaccgc cacgctgt
30001 tccaaccgcg acgggaagcc ggtgacatcc gggccgcgg cgatggacac cctggctcc
30061 cagctcaccc aaccggtgcg atgggacactg tgcacccgcga cgctgcgcga acacacatc
15 30121 acggcgatcg tggagttccc ccccgccggc acgcttagcg gtatcgccaa acgcaactt
30181 cgggggggttc cggcacgcgc cgtcaagtca cccgcagacc tggacgagct ggcaaaccta
30241 taaccgcgga ctggccaga acaaccatc acccgtagt tcgatttgcata cacaacatc
30301 tacgaaggga agcatgtgt gcctgtact caggaagaaa tcattgcgg tatcgccag
30361 atcatcgaag aggttaaccgg tatcgagccg tccgagatca ccccgagaa gtcgtcg
20 30421 gacgacccgtt acatcgactc gctgtcgatg gtcgagatcg ccgtgcagac cgaggacaag
30481 tacggcgtca agatccccga cgaggaccc gcccgtctgc gtaccgtcgg tgacgttgc
30541 gcctacatcc agaagctcgaa ggaagaaaac cggaggccgg ctcaggcggtt ggcgcgcaag
30601 attgagtcgg agaaccgcg tgccgttgcc aacgttcagg cgaggcttga ggccgagtcc
30661 aagtgagtca gccttccacc gctaattggcg gttcccgag cgttgcggatg accgcccgtca
25 30721 cagcgacgcgtc cggacatcg agagcacgtg gaagggtctg ttggccggcg
30781 agagcggcat ccacgcactc gaagacgagt tcgtcaccaaa gtggatcta gcggtcaaga
30841 tcggcgtca cctcaaggat cggcgacca gccacatggg ccgactcgac atgcgacgc
30901 tgtcgatgtt ccagcggatg ggcaagtgc tggcgccaca gctatgggag tccggccggca
30961 gcccggaggt cgatccagac cgggtcgccg ttgtgtcggtt caccggctta ggtggagccg
30 31021 agaggattgtt cgagagctac gacctgtatcg atgcggccgg ccccggaag gtgtccccgc
31081 tggccgttca gatgatcatcg cccaaacgggtt ccgcggccgtt gatcggtctg cagcttgggg
31141 cccgcgcggg ggtgtatgacc cgggtgtcggtt cctgttgc gggctcgaa ggcgtcgccc
31201 acgcgtggcg tcagatcgatcg atggcgacgc cgcacgtcgcc cgtctgcggc ggtgtcgaaag
31261 gaccatcgaa ggcgtgccc atcgccgggtt tctccatgtt gcccggcatcg tcgacccgca
35 31321 acgacgagcc tgagcggccccc tcccgccgt tcgacaagga cccgcacggc ttgtgttgc
31381 cggaggccgg tgcgtgtatcg tcacatcgaga cggaggagca cgcggaaagcc cgtggccca

31441 agccgttggc ccgattgctg ggtgccggta tcacctcgga cgccttcat atggggcgc
31501 ccgcggccga tgggttcgt gccggtaggg cgatgactcg ctcgctggag ctggccgggt
31561 tgcgcggcgc ggacatcgac cacgtcaacg cgcacggcac ggcacgcct atcgccgacg
5 31621 ccgcggaggc caacgccatc cgcgtcgccg gttgtatca ggccgggtg tacgcggcga
31681 agtctgcgt gggccactcg atcggcgcgg tcggtgctcg ctacgggtgc
31741 tgacgctgcg cgacggcgtc atcccgcga ccctgaacta cgagacaccc gatcccgaga
31801 tcgaccttga cgtcgtcgcc ggcaaccgc gctatggcga ttaccgtac gcagtcaca
31861 actcggttgcg gttcggcggc cacaatgtgg cgcttgcctt cggcggtac tgaagcacga
31921 catcgccggt cgcgaggccc gaggtggggg tcccccgct tgccggggcg agtcggaccg
10 31981 atatggaaagg aacgttcgca agaccaatga cggagctgg taccgggaaa gccttccct
32041 acgttagtcgt caccggcatac gccatgacga cgcgcgtcgc gaccgacgcg gagactacgt
32101 ggaagttgtt gctggaccgc caaagcggga tccgtacgct cgatgaccca ttctgcgagg
32161 agttcgcacct gccagttcgc atcggcggac atctgttga ggaattcgac caccagctga
15 32221 cgcggatcga actgcgcggg atggatacc tgcagcggat gtccaccgtg ctgagccggc
32281 gcctgtggga aaatgcggc tcacccgagg tggacaccaa tgcattgtg gtgtccatcg
32341 gcaccggcct gggttcggcc gaggaactgg tcttcagttt cgcgcgtatcg cgcgcgtcgc
32401 gaatgaaggc ggtctcgccg ctgaccgtgc agaagtacat gccaacggg gcccggcgg
32461 cggtcgggtt ggaacggcac gccaaggccg gggtgatgac gccggatcg gcgtgcgcatt
32521 cccgcgcga ggcacatcgcc cgtgcgtggc agcagatgt gctgggagag gccgatgcgg
20 32581 ccatctgcgg cggcgtggag accaggatcg aagcgggtcc catgcggggg ttgcgtcaga
32641 tgcgcatcgatgttccacc aacaacgcg accccgcggg tgcgcgcgc ccaattcgaca
32701 gggaccgcga cggctttgtt ttcggcgagg gcccgcctt tctgttgc tgcgttgc
32761 agcacgccaa ggcacgtggc gccaacatcc tggccggat catggcgcc agcatcacct
32821 ccgatggctt ccacatggtg gccccggacc ccaacggggg acgcgcggg catgcgatta
25 32881 cgcggcgat tcagctggcg ggcctcgccc cggcgacat cgaccacgtc aatgcgcacg
32941 ccacccggcac ccaggtcggc gacctggcg aaggcaggggc catcaacaac gcctggcg
33001 gcaacccgacc ggcgggttac gcccccaagt ctgcctcgcc ccactcggtt ggcgggtcg
33061 ggcgggttgcgatcgatctt acgggtctcg cttgcgcga tgcgttgc tgcgttgc
33121 tgaatctggtaaacctcgat cccgagatcg attggacgt ggtgggggt gacccgcac
33181 cgggcaatta cgggtatcgatcaataact cttgcgttgc ggcggccac aacgtggca
33241 tgcgcgttgcg acggtaactaa accccagcgat tgcgcgttgc gagacgttgc atgacaatca
33301 tggcccccga ggcgggttgcg gagtcgttgc acccccgca tccgttgc cggctgttgc
33361 actcttcgacgacggcgttgc tgcacgttgc tgaccgttgc ggagtgttgc
33421 cccggcgggg caccgtcaac ggtgttgc ccatcggtt ctgcaccgc ggcaccgttga
35 33481 tggccggcgcgatggcgatcgacgggttgc cgcacatcgatcaacgcctac gacactgcga

33541 tcgaagacca gagtcccatc gtggcatct ggcattcggg tgggtcccggtt ctggctgaag
 33601 gtgtcgccggc gtcacgcgtt gtaggcagg tggtcgaagc catgtccgc gctgtccggc
 33661 acatcccgca gatctcggtt gtcgtcggtt tcgcccggg cggccggcc tacggaccgg
 5 33721 cgttgaccga cgtcgtcgatc atggcgccgg aaagccgggtt gttcgaccc gggcccgacg
 33781 tgggtcgccag cgtcaccggc gaggacgtcg acatggcctc gtcgggtggg ccggagaccc
 33841 accacaagaa gtccgggtt tgccacatcg tcgcccacga cgaactcgat gcctacgacc
 33901 gtgggcggccg tgggtcgga ttgttctgcc agcaggggca ttgcgtcgac agcaaggccg
 33961 aggccgggtga caccgacatc cacgcgtc tgccgaatc ctgcgacgt gcctacgacg
 10 34021 tgcgtccgtatc cgtgacggccg atccgtatc cggacacacc gttcgacgag ttccaggcca
 34081 attgggcggcc gtcgatgggtt gtcgggtgg gtcggctgtc gggtcgcacg gtgggtgtac
 34141 tggccaacaa cccgctacgc ctggcggtt gcctgaactc cagaagcgca gagaaggcag
 34201 cgcgttcgtc gcccgggtgc gacgcgttcg ggattccgtt ggtgggtgt gtcgatgtgc
 34261 cggctatct gcccgggtgc gaccaggagt ggggtgggtt ggtgcgcgtt ggcgccaagt
 15 34321 tgcgtcacgc gttcgccgag tgccacgttc cgcgggtcac gtcgggtacc cagaagaccc
 34381 acggccggcc atacattcgatgtaactccc ggtcggttgcgaa cgcgaccaag gtgtcgccct
 34441 ggccggacgc cgagggtcgatc gtgatggcgatc aaggcgccatc ctgcacaaga
 34501 agaagttggc cgccgtcccg gagcacgaac gcaagcgctt gcaacgaccag ttggccggcc
 34561 agcatgagcg catcgccggc ggggtcgaca tgccgtcgatc gtcgacgaga
 20 34621 agatcgaccc ggcgcataact cgcagcaac tcaccgggc gtcggcgac gtcggccac
 34681 ggcgcggccg ccacaagaac atcccgctgt agttctgacc gcgacgacgc gcagaatcg
 34741 acgcgcgagg tccgcgttgc gcgatttcgc gtcgtcgatc cagttatccc cagcggtggc
 34801 tggcaacgc gaggcgctcc tcgcatgtc ggacgggtcc taccgacgcg ctaacaattc
 34861 tcgagaaggc cggcggttgc gcccaccacgc cgcattgtt caccgtcatg acccgccaa
 25 34921 agtcgcacgtt ccaagtgaaa aacggcgcc tcgttcgtt ttggtaacggg gtctacgcgg
 34981 cacaagagcc ggacctgttgc ggccgttgc cggcttcga tgcgttcatg gggggccacg
 35041 ccgtcggttgc tctggccacc gcccgggttgc tgcgttcatg gggggccacg
 35101 ctatccatat gtcgatccc ggagtaaggat gtcggccatc ggtcggttgc atggccacc
 35161 aacgcgtcgatc tgccggctcc caacgggttgc cagggtcgatc cgcgacgcg cccgcatgg
 30 35221 ctgcgttgc ggtcgacgc cagttcgcc gcccgggttgc gtcggccacc ctcgacgcgg
 35281 cactacggc aatcgctgc gtcgtcgatc aaattgaaa cgcgttgc gaggacgcg
 35341 gcccggagg catcggtcgatc ggcgcgttgc tgcgttcatg gggggccacg
 35401 cggccatggatc gaggcgaggctt cggcttcgtatc tgcgttcatg gggggccacg
 35461 ttcaataccatc gatacaccgcg caccgttgcgatc aaatgtggcg agtgcgttc gtcggccacc
 35521 acatcggttgc cgcggccgaa tacgaaagca tgcgttgc ggcggacccg gcccggatgc
 35581 tgcgttgc gacacgttgc gccaagctcc aagagcttgc gtcggacgattt gtcggccacc

35641 tcgtcgacga tgcagacgc gaacccggcc gcctggcgcc ccgcacatcgcc cgccacacctcg
35701 accgcgcgac tatggccggc tgaccgcgtgg tgagcagacg cagactcgca ctgcggccgg
35761 cgcagtgcga ctctgcgtct gctcgcgctc aacggctgag gaactccta gccacggcga
5 35821 ctacgcgtc gcgatcccgt ggcaccagac cgatccgggt ccggcggtcg aggtatcg
35881 ccacatccag cggcccccgtca tgggtcaccg cgtattcgaa ctccggccgg gtcacgtcga
35941 tgccgtcgcc gaccggctcg gtggccgct cacatgtggc ggccgcagcg acgtggccg
36001 cctcgcccccc gtaccgcgccc accagcgact cggcaatcc ggcgcggat ccggggggccg
36061 gcccagggtt cgccgggtcgcc cgatcagcg gcaggttgcg agtgcggcac ttgcggctc
10 36121 gcaggtgtcg cagcgtgatg ggcgcgttca gcacatcctc tgccatgttag cggtattccg
36181 tcagttgcc gccgaccaca ctgatcacgc cgcacggcga ttcaaaaaca gcgtggtcac
36241 gcgaaacgtc ggcgggtcgcc ccctggacac cagcaccgac ggtgtcgatt agcggccgca
36301 atcccgata ggcacccgtat acatccctgg tgccgaccgc cgtcccaat gcgtgttca
36361 ccgtatccag caggaacgtg atctcttccg aagacgggtt tggcacatcg ggaatcgggc
15 36421 cgggtgcgtc ttgcgtcggtc agcccgagat agatccggcc cagctgctcg ggcacatggcga
36481 acacgaagcg gttcagtcgatc cggggatcg gaatggtcag cgccgcagtc ggattggcaa
36541 acgacttcgc gtcgaagacc agatgtgtgc cgcggctggg gcgtagccctc agggacgggt
36601 cgcgttcacc cggccacacg cccggcggt tgatgcggc acgcgcgcac agcgcgaacg
36661 actgcccgggt ggcgggtcg gtcaactcca ccgaagtgcg ggtgacattc gacgcgccc
20 36721 cgtaaatgtgag gatgcggggcg ccgtgtggg ccgcgggtcgcc cgcgacggcc atgaccagcc
36781 gggcggtcgatc gatcaattgc ccgtgtacg cgagcagacc accgtcgagg cgtcccgcc
36841 gaacgggtggg agcaatctcc accacccgtg acgcggggat tcggcgatc cggggcaacg
36901 tcgcccggcg cgtacccgtc agcaccgcg aagcgtcgcc ggccaggaaa cggcacgc
36961 ccaacgccccg cttgggtgtga cccatcgacg gcaacaacgg gaccagttgc ggcacatggcat
25 37021 gcacgagatg aggagcgttg cgtgtcatca ggattccgcg ttgcacggcg ctgcggccgg
37081 cgcgtccac gttggcgctg gccagatgcg gcagaccgcg gtgcaccaac ttgcagctcc
37141 agcggctggt ggcgaacgcg agatcatgtt tttccaccaa ggccaccgtc agaccgggg
37201 tggcagcatc taaggcaatg ccaacaccgg taatgcgcgc gcctatcagc atgacgtcg
37261 gtgcggccacc gtcggccagt gcggtcgatc cggcgaggacg acgcgcgcg tttgagtcg
30 37321 ccgagtggtt catcagcaca aatatccgtt cagtcgtgg gtaagttcgg tggccagcgc
37381 ggcggaaatcg aggtcgaat cgacgtgc cgcggactgg atggtcgact gggcgatcag
37441 caacaccatg gtcgcgtc gacgagcgtc gccggagcgc acactgccc accgcgtcgc
37501 cactgtcagc cggcgccca acccctcgat caggacctgc tggctggtgc cgaggcgctc
37561 ggtgtatgtac accctggcca gtcggatgtt catgaccgcg atgatcagat cgtcacc
35 37621 caaccgggtcg gccacccgcgca caatctgcgtt taccacgc tcccggtcg ccccgatc
37681 gggcacctcc cgcagcgtc cggcgatatg gtcggatcg atggacgcgc tgatcgaccg

37741 ggtgtccggc cagcgacggt atacggtcgg gcggctcacg cccgcgcgcc gggcgatctc
 37801 ggcaagtgtc acccggtcca cgccgtaatc gacgacgcag ctgcgcgtg cccgcaggat
 37861 acgaccaccc gtatccgcgc ggtcattact cattgacagc atgtgtataa ctgtaacgcg
 5 37921 tgactcaccg cgaggaactc cttccaccga tgaatggga cgcgtggggat gatccgcgc
 37981 cggccaagcc actttctgtat ggcgtccggc cggtgctgaa gcagggtgt ggcctagcgg
 38041 actcggagca gcccgaactc gacccgcgc aggtgcagct gcccgcgtcc gcccgtcgg
 38101 gggcagacca (SEQ ID NO: 24)

10 6.9. X-linked Inhibitor of Apoptosis Protein (“XIAP”)
 GenBank Accession # U45880:

1 gaaaagggtgg acaaggccta tttcaagag aagatgactt ttaacagttt tgaaggatct
 61 aaaacttgtt tacctgcaga catcaataag gaagaagaat ttgtagaaga gtttaataga
 121 taaaaaactt ttgctaattt tccaagtggt agtcctgtt cagcatcaac actggcaca
 15 181 gcagggtttc ttatactgg tgaaggagat accgtgcggc gcttttagtt tcattgcagct
 241 gtagatagat ggcaatatgg agactcagca gttggaaagac acaggaaagt atccccaaat
 301 tgcagattta tcaacggctt ttatctgaa aatagtgcac cgcagtcac aaattctgg
 361 atccagaatg gtcagtcacaa agttgaaaac tatctggaa gcagagatca tttgccta
 421 gacaggccat ctgagacaca tgcagactat ctttgagaa ctggcaggt ttagatata
 481 tcagacacca tatacccgag gaaccctgcc atgtattgtt aagaagcttag attaaagtcc
 541 tttcagaact gcccagacta tgctcaccta accccaagag agttgcaag tgctggactc
 601 tactacacag gtattggtga ccaagtgcag tgctttgtt gtggggaaa actgaaaaat
 661 tgggaacctt gtgatcgtgc ctggcagaa cacaggcgcac actttctaa ttgcctttt
 721 gtttgggcc ggaatcttaa tattcgaagt gaatctgtat ctgtgagttc tgataggaat
 781 ttcccaaatt caacaatct tccaagaaat ccatccatgg cagattatga agcacggatc
 841 ttacatttg ggacatggat atactcagtt aacaaggaggc agcttgcaag agctggattt
 901 tatgttttag gtgaagggtga taaagttaag tgcttcaact gtggaggagg gctaactgat
 961 tggaagccca gtgaagaccc ttgggaacaa catgtaat ggtatccagg gtgcaaatat
 1021 ctgttagaac agaagggaca agaatataa aacaatattc attaactca ttacttgag
 1081 gagtgctgg taagaactac tgagaaaaca ccatcactaa ctagaagaat tgatgatacc
 1141 atcttccaaa atcctatggt acaagaagct atacgaatgg ggttcagttt caaggacatt
 1201 aagaaaataa tggaggaaaa aattcagata tctggagca actataatc acttgaggtt
 1261 ctgggtgcag atctgtgaa tgctcagaaa gacagtatgc aagatgagtc aagtcagact
 1321 tcattacaga aagagattag tactgaagag cagtaaggc gcctgcaaga ggagaagctt
 1381 tgcaaaaatct gtatggatag aaatattgtt atcggttttg ttccctgtgg acatctagtc
 35 1441 acttgtaaac aatgtgctga agcagttgac aagtgtccca tgcgtacac agtcattact

1501 ttcaagcaaa aaattttat gtcttaatct aactctatag taggcatgtt atgttgttct
 1561 tattaccctg attgaatgtg ttagtgtgaac tgactttaag taatcaggat tgaattccat
 1621 tagcatttc taccaagtag gaaaaaaaaat gtacatggca gtgttttagt tggcaatata
 1681 atcttgaat ttcttgattt tcagggat tagctgtatt atccattttt ttactgtta
 5 1741 tttaattgaa accatagact aagaataaga agcatcatac tataactgaa cacaatgtgt
 1801 attcatagta tactgatttta atttctaattt gtaagtgaat taatcatctg gattttttat
 1861 tctttcaga taggcttaac aaatggagct ttctgtatataatgtggag attagagta
 1921 atctccccaa tcacataattt tgtttgcgtt gaaaaaggaa taaaattgttc catgctggtg
 1981 gaaagataga gattgtttt agagggtgg tgggtgtt taggattctg tccattttct
 2041 tgtaaaggga taaacacgga cgtgtgcgaa atatgtttgtt aaagtgattt gccatttttg
 2101 aaagcgtatt taatgtataga atactatcga gccaacatgt actgacatgg aaagatgtca
 2161 gagatatgtt aagtgtaaaa tgcagaatggc gggacactat gtatgtctg agccagatca
 2221 aagtatgtat ttgtttaata tgcataagaac gagagattt gaaagatata caccaactg
 2281 ttaaatgtgg ttctcttcg gggagggggg gattggggg ggggccccag aggggtttta
 15 2341 gggggccctt ttcaacttcg actttttca ttgtgtctg ttggattttt ttataagtat
 2401 gtagaccccg aagggtttt tgggaactaa catcagtaac ctaaccccg tgactatcct
 2461 gtgtcttc tagggagctg tgggtttcc caccaccac cttccctct gaacaaatgc
 2521 ctgagtgcgtt gggcactttg (SEQ ID NO: 25)

20 General Target Region:

Internal Ribosome Entry Site (IRES) in 5' untranslated region:
 5'AGCUCCUAUAACAAAAGUCUGUUGCUUGUGUUUCACAUUUUGGAUU
 UCCUAAUAUAUAGUUCUCUUUUUAGAAAAGGUGGACAAGUCCUAUUU
 25 UCAAGAGAAG3' (SEQ ID NO: 26)

Initial Specific Target Motif:

RNP core binding site within XIAP IRES
 5'GGAUUUCUAAUUAUAGUUCUCUUUUU3' (SEQ ID NO: 27)

30 6.10. Survivin

GenBank Accession # NM_001168:

1 ccggccagatt tgaatgcgg gaccgggtgg cagagggtggc ggccggggca tgggtgcccc
 61 gacgttgccc cctgcctggc agccctttctt caaggaccac cgcattctta cattcaagaa
 35 121 ctggcccttc ttggagggct ggcgcgtcac cccggagcgg atggccgagg ctggcttcat
 181 ccactgcccc actgagaacg agccagactt ggcccaatgt ttctctgct tcaaggagct

241 ggaaggctgg gagccagatg acgacccat agaggaacat aaaaagcatt cgccgggtt
 301 cgcttcctt tctgtcaaga agcagttga agaattaacc ctggtaat ttgtaaaact
 361 ggacagagaa agagccaaga acaaattgc aaaggaaacc aacaataaga agaaagaatt
 5 421 tgaggaaact gcgaagaaag tgcgcgtgc catcgagcag ctggctgcca tggattgagg
 481 cctctggccg gagtcgcctg gtcccagagt ggctgcacca cttccagggt ttatccctg
 541 gtgccaccag cttcctgtg gccccttag caatgttta ggaaaggaga tcaacattt
 601 caaatttagat gttcaactg tgccttgtt ttgtctgaa agtggcacca gaggtgcitc
 661 tgcctgtca gcgggtgctg ctggtaacag tggctgctc tctctctc tctctttt
 721 ggggctcat ttgcgtt tgattcccg ggcttaccag gtgagaagtg agggaggaag
 10 781 aaggcagtgt ccctttgct agagctgaca gcttgcgtc cgtggcaga gcctccaca
 841 gtgaatgtgt ctggaccta tggatgtgag gctgtcacag tcctgagtgt ggacttggca
 901 ggtgcgtt gaatctgagc tgcagggtcc ttatctgtca caccgtgcc tcctcagagg
 961 acagttttt tggatgtgt tttttttt gtagatgca tgactgtgt
 1021 gtgatgagag aatggagaca gagtcctgg ctctctact gtttacaac atggttct
 1081 tattttgtt gaattgttaa ttcacagaat agcacaact acaattaaaa ctaagcacaa
 1141 agccattcta agtcattggg gaaacggggt gaactcagg tggatgagga gacagaatag
 1201 agtgatagga agcgtctggc agatactctt tttgcactg ctgtgtgatt agacaggccc
 1261 agtgagccgc gggcacatg ctggccgctc ctccctcaga aaaaggcagt ggcctaaatc
 1321 cttttaaat gactggcic gatgctgtt gggactggct gggctgtgc aggccgtgt
 1381 tctgtcagcc caacccac atctgtcac ttcacac gggggagaga cgcagtccgc
 1441 ccaggcccc gtttcttg gaggcagcag ctccgcagg gctgaagtct ggcttaagat
 1501 gatggatttgcattccctc ctccctgtca tagagctgca ggggtggatttgcatttc
 1561 gctggaaacc tctggaggc atctcggtt tccctgagaa ataaaaagcc tgcatttc (SEQ ID NO: 28)

25

**7. EXAMPLE: IDENTIFICATION OF A DYE-LABELED TARGET RNA
BOUND TO SMALL MOLECULAR WEIGHT COMPOUNDS**

The results presented in this Example indicate that gel mobility shift assays can be used to detect the binding of small molecules, such as the Tat peptide and 30 gentamicin, to their respective target RNAs.

7.1. Materials and Methods

7.1.1. Buffers

35 Tris-potassium chloride (TK) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100, and 0.5mM MgCl₂. Tris-borate-EDTA (TBE) buffer is

composed of 45 mM Tris-borate pH 8.0, and 1 mM EDTA. Tris-Potassium chloride-magnesium (TKM) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100 and 5mM MgCl₂.

5

7.1.1. Gel retardation analysis

RNA oligonucleotides were purchased from Dharmacon, Inc, Lafayette, CO). 500 pmole of either a 5' fluorescein labeled oligonucleotide corresponding to the 16S rRNA A site (5'-GGCGUCACACCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 29); Moazed & Noller, 1987, Nature 327:389-394; Woodcock *et al.*, 1991, EMBO J. 10:3099-3103; Yoshizawa *et al.*, 1998, EMBO J. 17:6437-6448) or a 5' fluorescein labeled oligonucleotide corresponding to the HIV-1 TAR element TAR RNA (5'-GGCGUCACACCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 30); Huq *et al.*, 1999, Nucleic Acids Research. 27:1084-1093; Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:12997-13002) was 3' labeled with 5'-³²P cytidine 3', 5'-bis(phosphate) (NEN) and T4 RNA ligase (NEBiolabs) in 10% DMSO as per manufacturer's instructions. The labeled oligonucleotides were purified using G-25 Sephadex columns (Boehringer Mannheim). For Tat-TAR gel retardation reactions the method of Huq *et al.* (Nucleic Acids Research, 1999, 27:1084-1093) was utilized with TK buffer containing 0.5mM MgCl₂ and a 12-mer Tat peptide (YGRKKRRQRRRP (SEQ ID NO: 31); single letter amino acid code). For 16S rRNA-gentamicin reactions, the method of Huq *et al.* was used with TKM buffer. In 20 μ l reaction volumes 50 pmoles of ³²P cytidine-labeled oligonucleotide and either gentamicin sulfate (Sigma) or the short Tat peptide (Tat₄₇₋₅₈) in TK or TKM buffer were heated at 90°C for 2 minutes and allow to cool to room temperature (approximately 24°C) over 2 hours. Then 10 μ l of 30% glycerol was added to each reaction tube and the entire sample was loaded onto a TBE non-denaturing polyacrylamide gel and electrophoresed at 1200-1600 volt-hours at 4°C. The gel was exposed to an intensifying screen and radioactivity was quantitated using a Typhoon phosphorimager (Molecular Dynamics).

30

7.2. Background

One method used to demonstrate small molecule interactions with natural occurring RNA structures such as ribosomes is by a method called chemical footprinting or toe printing (Moazed & Noller, 1987, Nature 327:389-394; Woodcock *et al.*, 1991, EMBO J. 10:3099-3103; Yoshizawa *et al.*, 1998, EMBO J. 17:6437-6448). Here the use of gel mobility shift assays to monitor RNA-small molecule interactions are described. This approach allows for rapid visualization of small molecule-RNA interactions based on the

difference between mobility of RNA alone versus RNA in a complex with a small molecule. To validate this approach, an RNA oligonucleotide corresponding to the well-characterized gentamicin binding site on the 16S rRNA (Moazed & Noller, 1987, *Nature* 327:389-394) and the equally well-characterized HIV-1 TAT protein binding site on the HIV-1 TAR element (Huq *et al.*, 1999, *Nucleic Acids Res.* 27: 1084-1093) were chosen. The purpose of these experiments is to lay the groundwork for the use of chromatographic techniques in a high throughput fashion, such as microcapillary electrophoresis, for drug discovery.

10

7.3. Results

A gel retardation assay was performed using the Tat₄₇₋₅₈ peptide and the TAR RNA oligonucleotide. As shown in Figure 1, in the presence of the Tat peptide, a clear shift is visible when the products are separated on a 12% non-denaturing polyacrylamide gel. In the reaction that lacks peptide, only the free RNA is visible. These observations confirm previous reports made using other Tat peptides (Hamy *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:3548-3553; Huq *et al.*, 1999, *Nucleic Acids Res.* 27: 1084-1093).

Based on the results of Figure 1, it was hypothesized that RNA interactions with small organic molecules could also be visualized using this method. As shown in Figure 2, the addition of varying concentrations of gentamicin to an RNA oligonucleotide corresponding to the 16S rRNA A site produces a mobility shift. These results demonstrate that the binding of the small molecule gentamicin to an RNA oligonucleotide having a defined structure in solution can be monitored using this approach. In addition, as shown in Figure 2, a concentration as low as 10ng/ml gentamicin produces the mobility shift.

To determine whether lower concentrations of gentamicin would be sufficient to produce a gel shift, similar experiment was performed, as shown in Figure 2, except that the concentrations of gentamicin ranged from 100 ng/ml to 10 pg/ml. As shown in Figure 3, gel mobility shifts are produced when the gentamicin concentration is as low as 10 pg/ml. Further, the results shown in Figure 3 demonstrate that the shift is specific to the 16S rRNA oligonucleotide as the use of an unrelated oligonucleotide, corresponding to the HIV TAR RNA element, does not result in a gel mobility shift when incubated with 10 μ g/ml gentamicin. In addition, if a concentration as low as 10 pg/ml gentamicin produces a gel mobility shift then it should be possible to detect changes to RNA structural motifs when small amounts of compound from a library of diverse compounds is screened in this fashion.

5 Further analysis of the gentamicin-RNA interaction indicates that the interaction is Mg²⁺- and temperature dependent. As shown in Figure 4, when MgCl₂ is not present (TK buffer), 1mg/ml of gentamicin must be added to the reaction to produce a gel shift.

10 Similarly, the temperature of the reaction when gentamicin is added is also important. When gentamicin is present in the reaction during the entire denaturation/renaturation cycle, that is, when gentamicin is added at 90°C or 85°C, a gel shift is visualized (data not shown). In contrast, when gentamicin is added after the renaturation step has proceeded to 75°C, a mobility shift is not produced. These results are 15 consistent with the notion that gentamicin may recognize and interact with an RNA structure formed early in the renaturation process.

20 8. EXAMPLE: **IDENTIFICATION OF A DYE-LABELED TARGET RNA
15 BOUNDED TO SMALL MOLECULAR WEIGHT COMPOUNDS
BY CAPILLARY ELECTROPHORESIS**

The results presented in this Example indicate that interactions between a peptide and its target RNA, such as the Tat peptide and TAR RNA, can be monitored by gel retardation assays in an automated capillary electrophoresis system.

25 8.1. **Materials and Methods**

30 8.1.1. **Buffers**

25 Tris-potassium chloride (TK) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100, and 0.5mM MgCl₂. Tris-borate-EDTA (TBE) buffer is composed of 45 mM Tris-borate pH 8.0, and 1 mM EDTA. Tris-Potassium chloride-magnesium (TKM) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100 and 5mM MgCl₂.

30 8.1.1. **Gel Retardation Analysis Using Capillary Electrophoresis**

35 RNA oligonucleotides were purchased from Dharmacon, Inc, Lafayette, CO). 500 pmole of a 5' fluorescein labeled oligonucleotide corresponding to the HIV-1 TAR element TAR RNA (5'-GGCGUCACACCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 30); Huq *et al.*, 1999, Nucleic Acids Research. 27:1084-1093; Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:12997-13002) was used. For Tat-TAR gel retardation reactions the method of Huq *et al.* (Nucleic Acids Research, 1999, 27:1084-1093) was

utilized with TK buffer containing 0.5mM MgCl₂ and a 12-mer Tat peptide (YGRKKRRQRRRP (SEQ ID NO: 31); single letter amino acid code). In 20 μ l reaction volumes 50 pmoles of labeled oligonucleotide and the short Tat peptide (Tat₄₇₋₅₈) in TK or 5 TKM buffer were heated at 90 °C for 2 minutes and allow to cool to room temperature (approximately 24 °C) over 2 hours. The reactions were loaded onto a SCE9610 automated capillary electrophoresis apparatus (SpectruMedix; State College, Pennsylvania).

8.2. Results

10 As presented in the previous Example in Section 7, interactions between a peptide and RNA can be monitored by gel retardation assays. It was hypothesized that interactions between a peptide and RNA could be monitored by gel retardation assays by an automated capillary electrophoresis system. To test this hypothesis, a gel retardation assay by an automated capillary electrophoresis system was performed using the Tat₄₇₋₅₈ peptide 15 and the TAR RNA oligonucleotide. As shown in Figure 5 using the capillary electrophoresis system, in the presence of the Tat peptide, a clear shift is visible upon the addition of increasing concentrations of Tat peptide. In the reaction that lacks peptide, only a peak corresponding to the free RNA is observed. These observations confirm previous reports made using other Tat peptides (Hamy *et al.*, 1997, Proc. Natl. Acad. Sci. USA 20 94:3548-3553; Huq *et al.*, 1999, Nucleic Acids Res. 27: 1084-1093).

25 The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

30 Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

35

The invention can be illustrated by the following embodiments enumerated in the numbered paragraphs that follow:

- 5 1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA: test compound complex is formed; (b) separating the detectably labeled target RNA: test compound complex formed in step(a) from uncomplexed target RNA molecules and test compounds; and (c) determining a structure of the test compound bound to the RNA in the RNA: test compound complex.
- 10 2. The method of paragraph 1 in which the target RNA molecule contains an HIV TAR element, internal ribosome entry site, "slippery site", instability element, or adenylate uridylate-rich element.
- 15 3. The method of paragraph 1 in which the RNA molecule is an element derived from the mRNA for tumor necrosis factor alpha ("TNF- α "), granulocyte-macrophage colony stimulating factor ("GM-CSF"), interleukin 2 ("IL-2"), interleukin 6 ("IL-6"), vascular endothelial growth factor ("VEGF"), human immunodeficiency virus I ("HIV-1"), hepatitis C virus ("HCV" - genotypes 1a & 1b), ribonuclease P RNA ("RNaseP"), X-linked inhibitor of apoptosis protein ("XIAP"), or survivin.
- 20 4. The method of paragraph 1 in which the detectably labeled RNA is labeled with a fluorescent dye, phosphorescent dye, ultraviolet dye, infrared dye, visible dye, radiolabel, enzyme, spectroscopic colorimetric label, affinity tag, or nanoparticle.
- 25 5. The method of paragraph 1 in which the test compound is selected from a combinatorial library comprising peptoids; random bio-oligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; and small organic molecule libraries, including but not limited to, libraries of benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.

6. The method of paragraph 1 in which screening a library of test compounds comprises contacting the test compound with the target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions.

5

7. The method of paragraph 6 in which the aqueous solution optionally further comprises non-specific nucleic acids comprising DNA, yeast tRNA, salmon sperm DNA, homoribopolymers, and nonspecific RNAs.

10

8. The method of paragraph 6 in which the aqueous solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. In another embodiment, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl₂. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂. In another embodiment, the solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant.

15

9. Any method that detects an altered physical property of a target nucleic acid complexed to a test compound from the unbound target nucleic acid may be used for separation of the complexed and non-complexed target nucleic acids in the method of paragraph 1. In a preferred embodiment, electrophoresis is used for separation of the complexed and non-complexed target nucleic acids. In a preferred embodiment, the electrophoresis is capillary electrophoresis. In other embodiments, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation are used for the separation of the complexed and non-complexed target nucleic acids.

20

25

10. The structure of the test compound of the RNA:test compound complex of paragraph 1 is determined, in part, by the type of library of test compounds. In a preferred embodiment wherein the combinatorial libraries are small organic molecule libraries, mass spectroscopy, NMR, or vibration spectroscopy are used to determine the structure of the test compounds.

30

35

WHAT IS CLAIMED IS:

1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of:
 - 5 (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA:test compound complex is formed;
 - 10 (b) separating the detectably labeled target RNA:test compound complex formed in step(a) from uncomplexed target RNA molecules and test compounds by capillary gel electrophoresis; and
 - 15 (c) determining a structure of the test compound bound to the RNA in the RNA:test compound complex by mass spectroscopy.

20

25

30

35

AMENDED CLAIMS

[received by the International Bureau on 17 September 2002 (17.09.02);
Claims 1 to 10 replaced by new claims 1 to 19. (3 sheets)]

5 1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of:

- 10 (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA: test compound complex is formed;
- (b) separating the detectably labeled target RNA: test compound complex formed in step (a) from uncomplexed target RNA molecules and test compounds; and
- 15 (c) determining a structure of the test compound bound to the RNA in the RNA: test compound complex.

2. The method of claim 1 in which the target RNA molecule contains an HIV TAR element, internal ribosome entry site, "slippery site", instability element, or
20 adenylate uridylate-rich element.

3. The method of claim 1 in which the RNA molecule is an element derived from the mRNA for tumor necrosis factor alpha ("TNF- α "), granulocyte-macrophage colony stimulating factor ("GM-CSF"), interleukin 2 ("IL-2"), interleukin 6
25 ("IL-6"), vascular endothelial growth factor ("VEGF"), human immunodeficiency virus I ("HIV-1"), hepatitis C virus ("HCV" - genotypes 1a & 1b), ribonuclease P RNA ("RNaseP"), X-linked inhibitor of apoptosis protein ("XIAP"), or survivin.

4. The method of claim 1 in which the detectably labeled RNA is
30 labeled with a fluorescent dye, phosphorescent dye, ultraviolet dye, infrared dye, visible dye, radiolabel, enzyme, spectroscopic colorimetric label, affinity tag, or nanoparticle.

5. The method of claim 1 in which the test compound is selected from a combinatorial library comprising peptoids; random bio-oligomers; diversomers such as
35 hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal

peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; or small organic molecule libraries.

5

6. The method of claim 5 in which the small organic molecule libraries are libraries of benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.

10

7. The method of claim 1 in which screening a library of test compounds comprises contacting the test compound with the target nucleic acid in the presence of an aqueous solution wherein the aqueous solution comprises a buffer and a combination of salts.

15

8. The method of claim 7 wherein the aqueous solution approximates or mimics physiologic conditions.

9. The method of claim 7 in which the aqueous solution optionally further comprises non-specific nucleic acids comprising DNA, yeast tRNA, salmon sperm DNA, homoribopolymers, and nonspecific RNAs.

10. The method of claim 7 in which the aqueous solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant.

25

11. The method of claim 10 in which the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl₂.

30

13 The method of claim 11 wherein the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂.

14. The method of claim 10 wherein the solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant.

35

15. The method of claim 1 in which separating the detectably labeled target RNA:test compound complex formed in step (a) from uncomplexed target RNA and 5 test compounds is by electrophoresis.

16. The method of claim 15 in which the electrophoresis is capillary electrophoresis.

10 17. The method of claim 1 in which separating the detectably labeled target RNA:test compound complex formed in step (a) from uncomplexed target RNA and test compounds is by fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, or 15 nanoparticle aggregation.

18. The method of claim 1 in which the library of test compounds are small organic molecule libraries.

20 19. The method of claim 18 in which the structure of the test compound is determined by mass spectroscopy, NMR, or vibration spectroscopy.

Figure 1
Sheet 1/5
Attorney Docket No. 10589-007

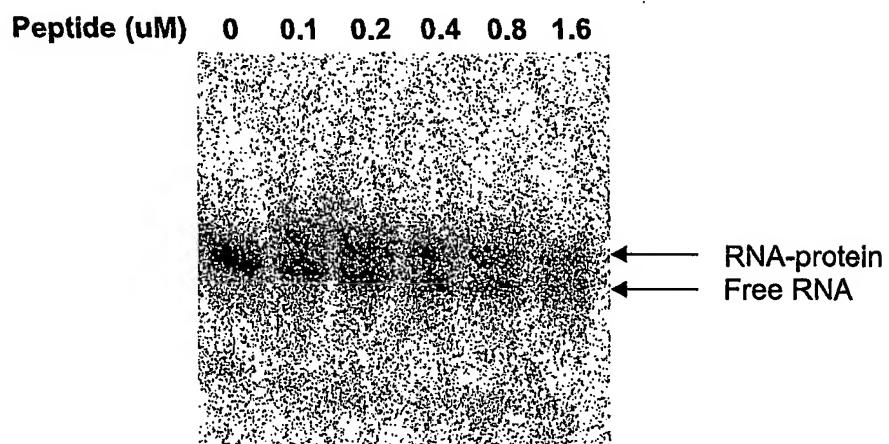


Figure 2
Sheet 2/5
Attorney Docket No. 10589-007

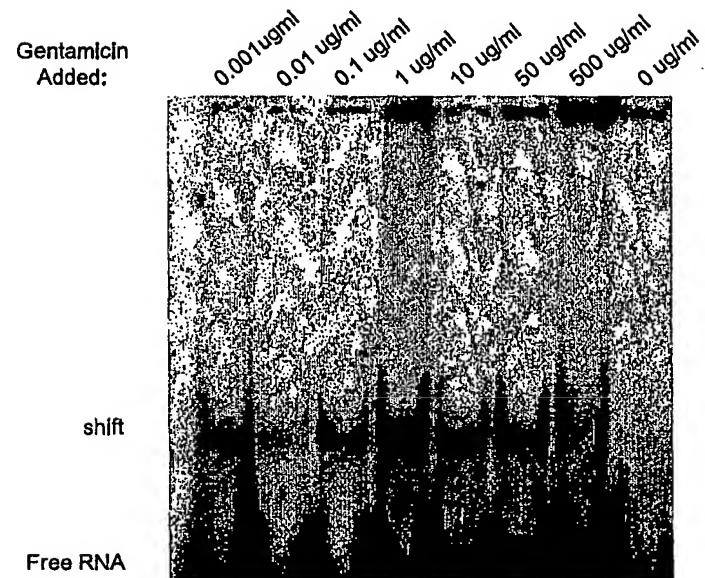


Figure 3
Sheet 3/5
Attorney Docket No. 10589-007

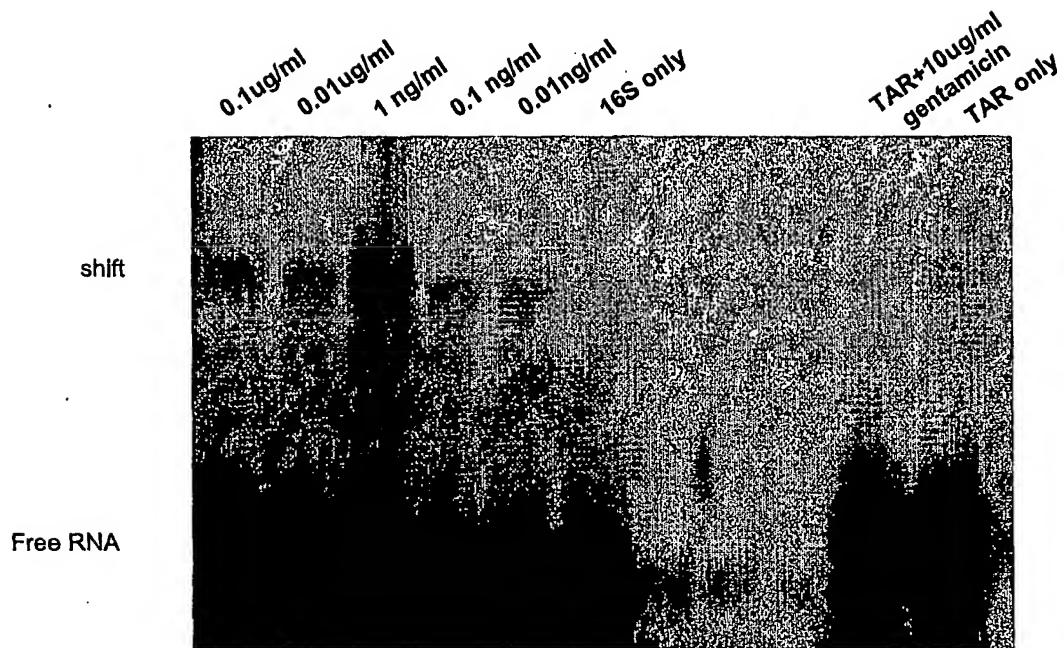


Figure 4
Sheet 4/5
Attorney Docket No. 10589-007

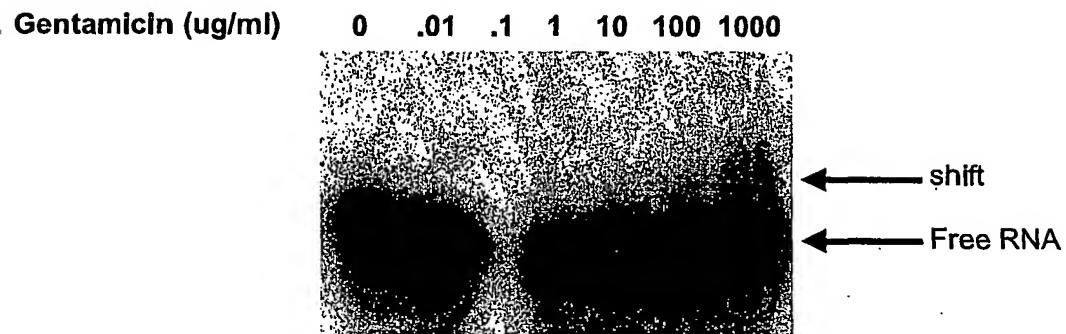
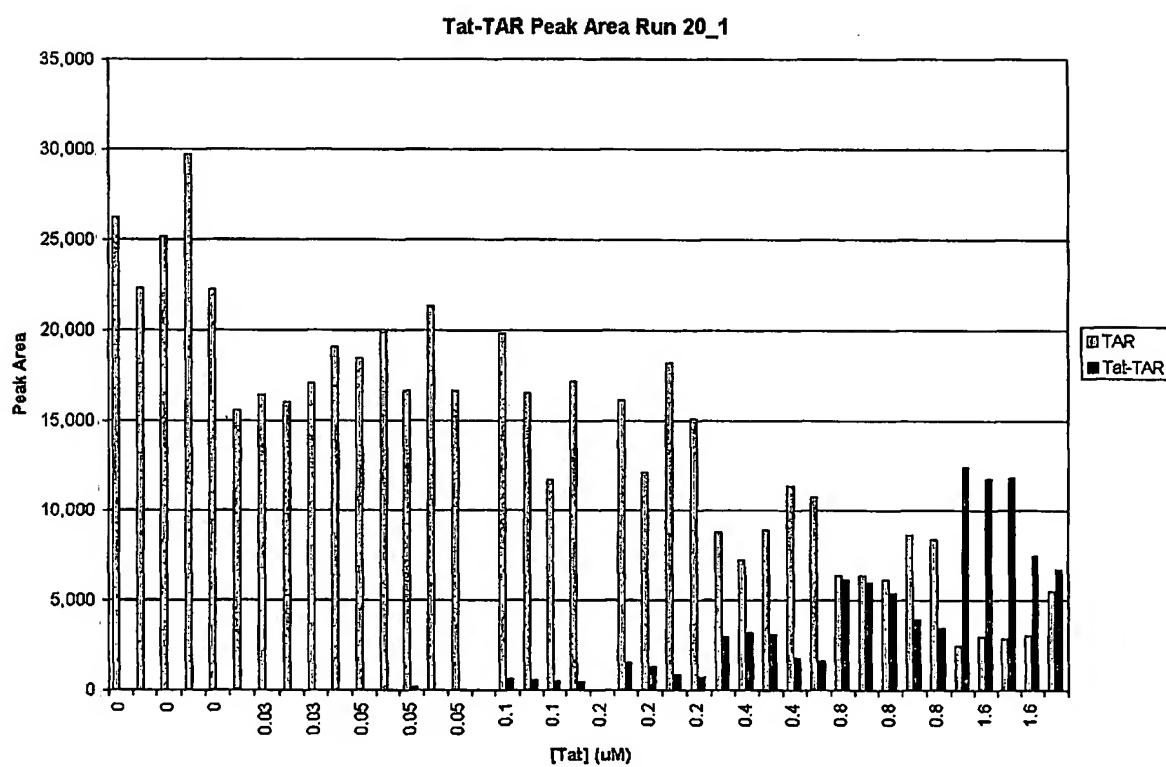


Figure 5
Sheet 5/5
Attorney Docket No. 10589-007



SEQUENCE LISTING

<110> PTC Therapeutics, Inc.

<120> METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA
STRUCTURAL MOTIFS

<130> 10589-007-228

<140> To be assigned

<141> 2002-04-11

<150> 60/282,965

<151> 2001-04-11

<160> 31

<170> PatentIn version 3.0

<210> 1

<211> 21

<212> RNA

<213> Homo sapiens

<400> 1
auuuauuuau uuauuuauuu a

21

<210> 2

<211> 17

<212> RNA

<213> Homo sapiens

<400> 2
auuuauuuau uuauuuua

17

<210> 3
<211> 15
<212> RNA
<213> Homo sapiens

<400> 3
wauuuauuuua uuuaw

15

<210> 4
<211> 13
<212> RNA
<213> Homo sapiens

<400> 4
wwauuuauuu aww

13

<210> 5
<211> 13
<212> RNA
<213> Homo sapiens

<400> 5
wwwauuuuaw www

13

<210> 6
<211> 1643
<212> DNA
<213> Homo sapiens

<400> 6
gcagaggacc agctaagagg gagagaagca actacagacc cccccctgaaa acaaccctca 60
gacgccacat cccctgacaa gctgccaggc aggttctttt cctctcacat actgaccac 120
ggctccaccc tctctcccct ggaaaggaca ccatgagcac tgaaagcatg atccgggacg 180
tggagctggc cgaggaggcg ctcccaaga agacaggggg gccccaggc tccaggcggt 240
gcttgttcct cagccttttc tccttcctga tcgtggcagg cgccaccacg ctttctgcc 300

tgctgcactt	tggagtgatc	ggcccccaga	ggaagagtt	ccccagggac	ctctctctaa	360
tcagccctct	ggcccaggca	gtcagatcat	cttctcgaac	cccgagtgac	aagcctgtag	420
cccatgttgt	agcaaaccct	caagctgagg	ggcagctcca	gtggctgaac	cgccgggcca	480
atgcctctct	ggccaatggc	gtggagctga	gagataacca	gctggtggtg	ccatcagagg	540
gcctgtacct	catctactcc	caggctct	tcaagggcca	aggctcccc	tcoacccatg	600
tgctcctcac	ccacaccatc	agccgcatcg	ccgtctccct	ccagaccaag	gtcaacctcc	660
tctctgccccat	caagagcccc	tgccagaggg	agacccaga	gggggctgag	gccaagccct	720
ggtatgagcc	catctatctg	ggaggggtct	tccagctgga	gaagggtgac	cgactcagcg	780
ctgagatcaa	tcggcccgac	tatctcgact	ttggccagtc	tggcaggtc	tactttggga	840
tcattggccct	gtgaggagga	cgaacatcca	accttcccaa	acgcctcccc	tgccccaaatc	900
cctttattac	cccctcccttc	agacaccctc	aacctcttct	ggctaaaaaa	gagaattggg	960
ggcttagggt	cggaacccaa	gcttagaact	ttaagcaaca	agaccaccac	ttcgaaacct	1020
gggattcagg	aatgtgtggc	ctgcacagtg	aattgctggc	aaccactaag	aattcaaact	1080
ggggcctcca	gaactcaactg	gggcctacag	cttgcattccc	tgacatctgg	aatctggaga	1140
ccagggagcc	tttggttctg	gccagaatgc	tgcaggactt	gagaagacct	cacctagaaaa	1200
ttgacacaag	tggaccttag	gccttcctct	ctccagatgt	ttccagactt	ccttgagaca	1260
cggagccca	ccctccccat	ggagccagct	ccctctattt	atgtttgcac	ttgtgattat	1320
ttattattta	tttattattt	atttatttac	agatgaatgt	atttatttgg	gagaccgggg	1380
tatcctgggg	gacccaatgt	aggagctgcc	ttggctcaga	catgtttcc	gtaaaaacgg	1440
agctgaacaa	taggctgttc	ccatgttagcc	ccctggcctc	tgtgccttct	tttgattatg	1500
ttttttaaaa	tatttatctg	attaagttgt	ctaaacaatg	ctgatttggt	gaccaactgt	1560
cactcattgc	tgagcctctg	ctccccaggg	gagttgtgtc	tgtaatcgcc	ctactattca	1620
gtggcgagaa	ataaagtttgc	ttt				1643

<210> 7

<211> 756

<212> DNA

<213> Homo sapiens

<400> 7	gctggaggat	gtggctgcag	agcctgctgc	tcttgggcac	tgtggcctgc	agcatctctg	60
	cacccgcccc	ctcgccccagc	cccagcacgc	agccctggga	gcatgtgaat	gccatccagg	120
	aggcccgccg	tctcctgaac	ctgagtagag	acactgctgc	tgagatgaat	gaaacagtag	180

aagtcatctc agaaatgttt	240
gacctccagg agccgacctg	
cctacagacc cgccctggagc	
tgtacaagca gggcctgcgg ggcagcctca	300
ccaagctcaa gggcccttg accatgatgg	
ccagccacta caagcagcac tgccctccaa	360
ccccggaaac ttccctgtgca acccagacta	
tcaccttga aagtttcaaa gagaacctga	420
aggactttct gcttgcata cccttgact	
gctgggagcc agtccaggag tgagacggc	480
cagatgaggc tggccaagcc ggggagctgc	
tctctcatga aacaagagct agaaactcag	540
gatggtcata ttggaggac caaggggtgg	
gccacagcca tggtgggagt ggcctggacc	600
tgccctgggc cacactgacc ctgatacagg	
catggcagaa gaatgggaat attttatact	660
gacagaaatc agtaatattt atatatttat	
atttttaaaa tatttattta tttattttata	720
taagttcata ttccatattt attcaagatg	
ttttaccgta ataattatta taaaaatat gcttct	756

<210> 8

<211> 756

<212> DNA

<213> Homo sapiens

<400> 8	
tctggaggat gtggctgcag	60
agcctgctgc tcttggcac tggcctgc agcatctcg	
cacccggcccg ctcgcccagc	120
cccagcacgc agccctggga	
gcatgtaat gccatccagg	
aggcccgccg tctcctgaac	180
ctgagtagag acactgctgc	
tgagatgaat gaaacagtag	
aagtcatctc agaaatgttt	240
gacctccagg agccgacctg	
cctacagacc cgccctggagc	
tgtacaagca gggcctgcgg	300
ggcagcctca ccaagctcaa	
gggcccttg accatgatgg	
ccagccacta caagcagcac	360
tgccctccaa cccccggaaac	
ttccctgtgca acccagacta	
tcaccttga aagtttcaaa	420
gagaacctga aggactttct	
gcttgcata cccttgact	
gctgggagcc agtccaggag	480
tgagacggc cagatgaggc	
tggccaagcc ggggagctgc	
tctctcatga aacaagagct	540
agaaactcag gatggtcata	
ttggaggac caaggggtgg	
gccacagcca tggtgggagt	600
ggcctggacc tgccctgggc	
cacactgacc ctgatacagg	
catggcagaa gaatgggaat	660
attttatact gacagaaatc	
agtaatattt atatatttat	
atttttaaaa tatttattta	720
tttattttata taagttcata	
ttccatattt attcaagatg	
ttttaccgta ataattatta	756
taaaaatat gcttct	

<210> 9

<211> 825

<212> DNA

<213> Homo sapiens

<400> 9
atcactctct ttaatcacta ctcacattaa cctcaactcc tgccacaatg tacaggatgc 60
aactcctgtc ttgcattgca ctaattcttg cacttgcac aaacagtgca cctacttcaa 120
gttcgacaaa gaaaacacaaag aaaacacagc tacaactgga gcatttactg ctggatttac 180
agatgatttt gaatggaatt aataattaca agaatccaa actcaccagg atgctcacat 240
ttaagttta catgccaaag aaggccacag aactgaaaca gtttcagtgt ctagaagaag 300
aactcaaacc tctggaggaa gtgctgaatt tagctcaaag caaaaacttt cacttaagac 360
ccagggactt aatcagcaat atcaacgtaa tagttctgga actaaaggaa tctgaaacaa 420
cattcatgtg tgaatatgca gatgagacag caaccattgt agaatttctg aacagatgga 480
ttaccttttgc tcaaagcattc atctcaacac taacttgata attaagtgt tcccaacttaa 540
aacatatcag gccttctatt tatttattta aatatttaaa ttttatattt attgttgaat 600
gtatggttgc tacctattgt aactattatt ctaatctta aaactataaa tatggatctt 660
ttatgattct ttttgttaagc cctagggct ctaaaatggt ttaccttatt tatccaaaa 720
atatttatta ttatggttcaa tgttaaatat agtatctatg tagattggtt agtaaaaacta 780
tttaataaaat ttgataaaata taaaaaaaaaaa aaacaaaaaaaaaa aaaaaa 825

<210> 10

<211> 15

<212> RNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)..(1)

<223> N = A, U, G, OR C

<220>

<221> misc_feature

<222> (15)..(15)

<223> N = A, U, G, OR C

<400> 10
nauuuauuuua uuuuau 15

<210> 11

<211> 1125

<212> DNA

<213> Homo sapiens

<400> 11
ttctgcctc gagcccaccc ggaacgaaag agaagctcta tctgcctcc aggagcccag 60
ctatgaactc cttctccaca agcgccctcg gtccagttgc cttctccctg gggctgtcc 120
tggtgttgc tgctgccttc cctgccccag tacccccagg agaagattcc aaagatgttag 180
ccgccccaca cagacagcca ctcacccctt cagaacgaat tgacaaacaa attcggtaca 240
tcctcgacgg catctcagcc ctgagaaagg agacatgtaa caagagtaac atgtgtgaaa 300
gcagcaaaga ggcactggca gaaaacaacc tgaaccccttcc aaagatggct gaaaaagatg 360
gatgtttcca atctggattc aatgaggaga cttgcctggt gaaaatcatc actggcttt 420
tggagtttga ggtataccta ggtaccccttcc agaacagatt tgagagtagt gaggaacaag 480
ccagagctgt gcagatgagt acaaaggatcc tgcacccatc cctgcagaaaa aaggccaaaga 540
atctagatgc aataaccacc cctgacccaa ccacaaatgc cagcctgctg acgaagctgc 600
aggcacagaa ccagtggctg caggacatga caactcatct cattctgcgc agctttaagg 660
agttcctgca gtccagccctg agggctcttc ggcaaatgtt gcatggcac ctcagattgt 720
tgggtttaat gggcattccct tcttctggtc agaaacctgt ccactggca cagaacttat 780
gttggctct atggagaact aaaagtatga gcgttaggac actatttaa ttatttttaa 840
tttattaata ttaaatatg tgaagctgag ttaatttatg taagtcatat ttatatttt 900
aagaagtacc acttggaaaca ttttatgtat tagttttgaa ataataatgg aaagtggcta 960
tgcagtttga atatcctttt tttcagagcc agatcatttc ttggaaagtg taggcttacc 1020
tcaaataaat ggctaactta tacatattt taaagaaata tttatattgt atttatataa 1080
tgtataaatg gtttttatac caataaatgg cattttaaaa aattc 1125

<210> 12

<211> 3166

<212> DNA

<213> Homo sapiens

<400> 12
 aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagac gcgcgggcgt 60
 gcgagcagcg aaagcgacag gggcaaagtg agtgacctgc ttttgggggt gaccgcccga 120
 gcgccgcgtg agccctcccc cttggatcc cgca gctgac cagtcgcgt gacggacaga 180
 cagacagaca cccgccccag cccca gttac cacccctcc cccggccggcg gcggacagtg 240
 gacgcggcgg cgagccgcgg gcaggggccc gagcccgccc cccggaggcgg ggtggagggg 300
 gtcggagctc gcggcgtcgc actgaaactt ttctgtccaaac ttctggctg ttctcgcttc 360
 ggaggagccg tggtccgcgc gggggaaagcc gagccgagcg gagccgcgag aagtgcgtac 420
 tcggccggg aggagccgca gcccggaggag ggggaggagg aagaagagaa ggaagaggag 480
 agggggccgc agtggcgact cggcgctcgg aagccgggct catggacggg tgaggcggcg 540
 gtgtgcgcag acagtgcgtcc agcgcgcgcg ctccccagcc ctggccggc ctcggccgg 600
 gaggaagagt agctcgccga ggcgcgagg agagcgggccc gccccacagc ccgagccgga 660
 gagggacgcg agccgcgcgc cccggcggg cctccgaaac catgaacttt ctgtgtctt 720
 ggggtgcattt gggcatttgc ttgtgtctt acctccacca tgccaaatgg tccaggctg 780
 caccatggc agaaggaggaa gggcagaatc atcacgaatg ggtgaatgtc atggatgtct 840
 atcagcgccag ctactgcattt ccaatcgaga ccctgggtgaa catcttccag gagtaccctg 900
 atgagatcga gtacatcttc aagccatcctt gtgtgtccct gatgcgtgc gggggctgtc 960
 ccaatgacga gggcctggag tgggtgcctt ctgaggagtc caacatcacc atgcagatta 1020
 tgccggatcaa acctcaccaa ggccagcaca taggagagat gagcttccta cagcacaaca 1080
 aatgtgaatg cagacccaaag aaagatagag caagacaaga aaatccctgt gggcattgtc 1140
 cagagcggag aaagcatttgg tttgtacaag atccgcagac gtgtaaatgt tcctgcaaaa 1200
 acacacactc gcgttgcag ggcaggcagc ttgagttaaa cgaacgtact tgcagatgt 1260
 acaagccgag gcggtgagcc gggcaggagg aaggagcctc cctcagggtt tcgggaacca 1320
 gatctcttc caggaaagac tgatacagaa cgtatcgatac agaaaccacg ctggcccac 1380
 cacaccatca ccatcgacag aacagtccctt aatccagaaaa cctgaaatga aggaagagga 1440
 gactctgcgc agagcacttt gggtccggag ggcgagactc cggcggaaacg attccgggc 1500
 ggggtgacccca gcacggtcccc ttttggattt ggattcgcca ttttattttt cttgtgtcta 1560
 aatcaccgag cccggaaagat tagagatgtt tatttctggg attcctgttag acacacccac 1620
 ccacatacat acatttat atatatatat tatatatata taaaaataaa tatctctatt 1680
 ttatatatat aaaatatata tattttttt ttaaattaac agtgcataatg ttattgggtgt 1740
 cttcactgga tggatggac tgctgtggac ttgagttggg agggaaatgt tccactcag 1800

atcctgacag ggaagaggag gagatgagag actctggcat gatcttttt ttgtcccaact	1860
tggtggggcc agggtcctct cccctgccc aagaatgtgca aggccaggc atgggggcaa	1920
atatgaccca gttttggaa caccgacaaa cccagccctg gcgctgagcc tctctacccc	1980
aggtcagacg gacagaaaga caaatcacag gttccggat gaggacaccg gctctgacca	2040
ggagtttggg gagcttcagg acattgctgt gctttggga ttccctccac atgctgcacg	2100
cgcacatctcg ccccaaggc actgcctgga agattcagga gcctggcgg ccttcgctta	2160
ctctcacctg cttctgagtt gcccaggagg ccactggcag atgtcccgcc gaagagaaga	2220
gacacattgt tggagaaga agcccatgac agcgccccctt cctggactc gccctcatcc	2280
tcttcctgct ccccttcctg gggtgcagcc taaaaggacc tatgtcctca caccattgaa	2340
accactagtt ctgtcccccc aggaaacctg gttgtgtgtg tgtgagtgg tggacttcct	2400
ccatccccctg gtccttcct tccctcccg aggcacagag agacaggca ggatccacgt	2460
gcccattgtg gaggcagaga aaagagaaag tgtttatat acggtaactt ttaatatatcc	2520
cttttaatt agaaattaga acagtttaatt taattaaaga gtagggttt tttcagttat	2580
tcttggtaa tatttaattt caactattta tgagatgtat ctttgctct ctttgctct	2640
cttatttgta ccggttttt tatataaaat tcatgtttcc aatctcttc tccctgatcg	2700
gtgacagtca ctatgttatac ttgaacagat atttaattt gctaacaactc agctctgccc	2760
tccccgatcc cctggctccc cagcacacat tccttgaaa gagggttca atatacatct	2820
acatactata tatatattgg gcaacttcta ttgtgtgtat tatataatata tatatgtta	2880
tgtatatatg tgatcctgaa aaaataaaca tcgctattct gtttttata tgttcaaacc	2940
aaacaagaaa aaatagagaa ttctacatac taaaatcttc tccttttta attttatat	3000
ttgttatcat ttatattattg gtgctactgt ttatccgtaa taattgtggg gaaaagatat	3060
taacatcagc tctttgtctc tagtgcagg tttcgagata ttccgtagta catatttt	3120
tttaaacaac gacaaagaaa tacagatata tcttaaaaaa aaaaaa	3166

<210> 13

<211> 249

<212> RNA

<213> Homo sapiens

<400> 13	
ccggcucau ggacggguga ggccggcggug ugcgcagaca gugcucccagc gcgcgccuc	60
cccaagccug gcccggccuc gggccgggag gaagaguagc ucggccgaggc gccgaggaga	120
gcgggcccgc ccacagcccg agccggagag ggacgcgagc cgccgcggccc ggucggccu	180

ccgaaaccau	gaacuuucug	cugucuuggg	ugcauuggag	ccuugccuug	cugcucuacc	240
uccaccaug						249

<210> 14

<211> 9181

<212> DNA

<213> Homo sapiens

<400> 14						
ggtctctctg	gttagaccag	atctgagcct	gggagctctc	tggctaacta	gggaacccac	60
tgcttaagcc	tcaataaagc	ttgccttgag	tgcttcaagt	agtgtgtgcc	cgtctgttgt	120
gtgactctgg	taactagaga	tccctcagac	ccttttagtc	agtgtggaaa	atctctagca	180
gtggcgcccg	aacagggacc	tgaaagcgaa	aggaaacca	gaggagctct	ctcgacgcag	240
gactcggctt	gctgaagcgc	gcacggcaag	aggcgagggg	cggcgactgg	ttagtacgcc	300
aaaaattttg	actagcgag	gctagaagga	gagagatggg	tgcgagagcg	tcagtattaa	360
gcgggggaga	attagatcga	tgggaaaaaa	ttcggttaag	gccaggggga	aagaaaaaat	420
ataaaattaaa	acatatagtt	tgggcaagca	gggagctaga	acgattcgca	gttaatcctg	480
gcctgttaga	aacatcgaaa	ggctgttagac	aaatactggg	acagctacaa	ccatcccttc	540
agacaggatc	agaagaactt	agatcattat	ataatacagt	agcaaccctc	tattgtgtgc	600
atcaaaggat	agagataaaa	gacaccaagg	aagctttaga	caagatagag	gaagagcaaa	660
acaaaagtaa	aaaaaaagca	cagcaagcag	cagctgacac	aggacacagc	aatcaggtca	720
gccaaaatta	ccctatagtg	cagaacatcc	aggggcaaat	ggtacatcag	gccatatcac	780
ctagaacttt	aatgcatgg	gtaaaagtag	tagaagagaa	ggctttcagc	ccagaagtga	840
tacccatgtt	ttcagcatta	tcagaaggag	ccacccacaa	agatttaaac	accatgctaa	900
acacagtggg	gggacatcaa	gcagccatgc	aaatgttaaa	agagaccatc	aatgaggaag	960
ctgcagaatg	ggatagagtg	catccagtgc	atgcagggcc	tattgcacca	ggccagatga	1020
gagaaccaag	gggaagtgac	atagcaggaa	ctactagtac	cttcaggaa	caaataggat	1080
ggatgacaaa	taatccaccc	atcccagttag	gagaaattta	taaaagatgg	ataatcctgg	1140
gattaaataa	aatagtaaga	atgtatagcc	ctaccagcat	tctggacata	agacaaggac	1200
caaaggaacc	cttttagagac	tatgttagacc	ggttctataa	aactctaaga	gccgagcaag	1260
cttcacagga	gtaaaaaaaaat	tggatgacag	aaaccttgtt	ggtccaaaat	gcgaacccag	1320
attgttaagac	tattttaaaaa	gcattggac	cagcggctac	actagaagaa	atgtatgacag	1380
catgtcaggg	agtaggagga	cccgccata	aggcaagat	tttggctgaa	gcaatgagcc	1440

aagtaacaaa	ttcagctacc	ataatgatgc	agagaggcaa	ttttaggaac	caaagaaaaga	1500
ttgttaagtg	tttcaattgt	ggcaaagaag	ggcacacagc	cagaaattgc	aggccccta	1560
ggaaaaagg	ctgttgaaa	tgtggaaagg	aaggacacca	aatgaaagat	tgtactgaga	1620
gacaggctaa	tttttaggg	aagatctggc	cttcctacaa	gggaaggcca	gggaattttc	1680
ttcagagcag	accagagcca	acagccccac	cagaagagag	cttcaggctc	gggtagaga	1740
caacaactcc	ccctcagaag	caggagccga	tagacaagga	actgtatcct	ttaacttccc	1800
tcaggtca	cttggcaac	gaccctcg	cacaataaag	atagggggc	aactaaagga	1860
agctctatta	gatacaggag	cagatgatac	agtattagaa	gaaatgagtt	tgccaggaag	1920
atggaaacca	aaaatgata	ggggattgg	agtttttac	aaagtaagac	agtatgatca	1980
gatactcata	gaaatctgt	gacataaagc	tataggtaca	gtatttagtag	gacctacacc	2040
tgtcaacata	attggaagaa	atctgttgc	tcagattggt	tgcactttaa	atttcccat	2100
tagccattt	gagactgtac	cagtaaaatt	aaagccagga	atggatggcc	caaaagttaa	2160
acaatggcca	ttgacagaag	aaaaaataaa	agcattagta	gaaatttga	cagagatgga	2220
aaaggaaggg	aaaatttcaa	aaattggcc	tgaaaatcca	tacaatactc	cagtatttgc	2280
cataaaagaaa	aaagacagta	ctaaatggag	aaaattagta	gatttcagag	aacttaataa	2340
gagaactcaa	gacttctgg	aagttcaatt	aggaatacc	catccgcag	ggtaaaaaaaa	2400
gaaaaaatca	gtaacagtac	tggatgtgg	tgatgcata	ttttcagttc	ccttagatga	2460
agacttcagg	aagtatactg	catttaccat	acctagtata	aacaatgaga	caccagggat	2520
tagatatcag	tacaatgtgc	ttccacaggg	atggaaagga	tcaccagcaa	tattccaaag	2580
tagcatgaca	aaaatcttag	agccttttag	aaaacaaaat	ccagacatag	ttatctatca	2640
atacatggat	gatttgtatg	taggatctga	cttagaaata	ggcagcata	gaacaaaaat	2700
agaggagctg	agacaacatc	tgttgaggt	gggacttacc	acaccagaca	aaaaacatca	2760
gaaagaacct	ccattcctt	gatgggta	tgaactccat	cctgataaat	ggacagtaca	2820
gcctatagt	ctgccagaaa	aagacagctg	gactgtcaat	gacatacaga	agtttagtgg	2880
gaaattgaat	tggcaagtc	agatttaccc	aggattaaa	gtaaggcaat	tatgtaaact	2940
ccttagagga	accaaagcac	taacagaagt	aataccacta	acagaagaag	cagagctaga	3000
actggcagaa	aacagagaga	ttctaaaaga	accagtacat	ggagtgtatt	atgaccatc	3060
aaaagactta	atagcagaaa	tacagaagca	ggggcaaggc	caatggacat	atcaaattta	3120
tcaagagcca	tttaaaaatc	tgaaaacagg	aaaatatgca	agaatgaggg	gtgcccacac	3180
taatgatgta	aaacaattaa	cagaggcagt	gcaaaaaata	accacagaaa	gcatacata	3240
atggggaaag	actcctaaat	ttaaactgcc	cataaaaaag	gaaacatggg	aaacatggt	3300
gacagagtat	tggcaagcca	cctggattcc	tgagtggag	tttgttata	ccctccctt	3360

agtgaaat~~ta~~ tggtaccagt tagagaaaga acccatagta ggagcagaaa ccttctatgt 3420
agatgggc~~a~~ gctaacaggg agactaaatt agaaaaagca ggatatgtta ctaatagagg 3480
aagacaaaaa gttgtcacc~~c~~ taactgacac aacaaat~~c~~ aagactgagt tacaagcaat 3540
ttatctagct ttgcaggatt cg~~g~~gattaga agtaaacata gtaacagact cacaat~~at~~gc 3600
attaggaat~~c~~ attcaagcac aaccagat~~c~~ aagtgaat~~c~~ gagtt~~g~~at~~c~~ atcaaataat 3660
agagcagtta ataaaaaagg aaaagg~~t~~cta tctggcatgg gtaccagcac acaaaggaat 3720
tggag~~g~~aat gaacaagtag ataaatt~~g~~t cagt~~g~~ct~~g~~ga atcaggaaag tactat~~ttt~~ 3780
agatggaata gataaggccc aagatgaaca tgagaaat~~at~~ cacagtaatt ggagagcaat 3840
ggct~~g~~at~~g~~tat tttaac~~c~~tc~~g~~ cac~~t~~gtagt agcaaaagaa atagtagcc~~a~~ gctgtgataa 3900
atgtcagcta aaaggagaag ccatgcatgg acaagtagac t~~g~~tagtccag gaat~~at~~ggca 3960
actagatt~~g~~t acacat~~t~~tag aaggaaaagt tatcctggta gcagttcatg tagccagtgg 4020
atata~~t~~agaa gcagaagtt~~a~~ ttccagcaga aacagggcag gaaacagcat attttctttt 4080
aaaatt~~g~~ca ggaagatggc cagtaaaaac aatacatact gacaatggca gcaat~~t~~tcac 4140
cggtgctac~~g~~ gtagggc~~g~~ cctgttgg~~t~~ ggcgggaat~~c~~ aagcaggaat ttggaattcc 4200
ctacaatccc caaagtcaag gagtagtaga atctat~~g~~aa~~t~~ aaagaattaa agaaaattat 4260
aggacag~~g~~ta agagatcagg ctgaacat~~c~~ taagacagca gtacaatgg cag~~t~~attcat 4320
ccacaat~~ttt~~ aaaagaaaag gggggattgg ggggtacagt gcagggaaa gaat~~g~~ttag~~t~~aga 4380
cataat~~g~~ca acagacatac aaactaaaga attacaaaaa caaattacaa aaattcaaaaa 4440
ttt~~c~~gggtt tattacaggg acagcagaaa tccactttgg aaaggaccag caaagctcct 4500
ctggaaaggt gaaggggcag tagtaatac~~a~~ agataat~~g~~t gacataaaag tagtgcc~~a~~ag 4560
aagaaaagca aagatcatta gggattatgg aaaacagat~~g~~ gcaggtgat~~g~~ attgtgtggc 4620
aagt~~g~~acag gatgaggatt agaacatgg aaagtttagt aaaacaccat atgtatgttt 4680
cagggaaagc tagggatgg tttat~~g~~ac atcactat~~g~~a aagccctcat ccaagaataa 4740
gtt~~c~~agaagt acacatccc~~a~~ ctagggat~~g~~ ctagattgg~~t~~ aataacaaca tattggg~~g~~tc 4800
tgcatacagg agaaagagac tggcatttgg gtcagggagt ctccat~~g~~aa tggaggaaaa 4860
agagat~~g~~at~~g~~ cacacaagta gaccctgaac tagcagacca actaattcat ctgtattact 4920
ttgactgttt ttcagact~~c~~ gctataagaa aggcc~~t~~tatt aggacacata gttagcccta 4980
gg~~t~~gtgaata tcaagcagga cataacaagg tag~~g~~at~~c~~t~~c~~ acaatact~~g~~ gcactagcag 5040
cattaataac accaaaaaaag ataaagccac ctttgcctag tg~~t~~acgaaa ctgacagagg 5100
atagatggaa caagccccag aagaccaagg gcccacagagg gagccacaca atgaatggac 5160
actagagctt ttagaggagc ttaagaat~~g~~a agctgttaga cattttccta ggatttggct 5220
ccatggctta gggcaacata tctat~~g~~aaac ttatgggat acttggcag gagtggaagc 5280

cataataaga attctgcaac aactgctgtt tatccatttt cagaattggg tgcgacata	5340
gcagaatagg cgttactcga cagaggagag caagaaatgg agccagtaga tcctagacta	5400
gagccctgga agcatccagg aagtcagcct aaaactgctt gtaccaattt ctattgtaaa	5460
aagtgttgct ttcattgcca agtttggcataa acacaaaag ccttaggcattcctatggc	5520
aggaagaagc ggagacagcg acgaagagct catcagaaca gtcagactca tcaagcttct	5580
ctatcaaagc agtaagtatg acatgtatg caacctatac caatagtagc aatagtagca	5640
tttagtagtag caataataat agcaatagtt gtgtggtcca tagtaatcat agaatatagg	5700
aaaatattaa gacaaagaaa aatagacagg ttaattgata gactaataga aagagcagaa	5760
gacagtggca atgagagtga aggagaaata tcagcacttg tggagatggg ggtggagatg	5820
gggcaccatg ctccttggga tggatgtatgat ctgtgtgtt acagaaaaat tggggtcac	5880
agtctattat ggggtacctg tgtgaaagga agcaaccacc actctatttt gtgcattcaga	5940
tgctaaagca tatgatacag aggtacataa tggggcc acacatgcct gtgtacccac	6000
agaccccaac ccacaagaag tagtattggt aaatgtgaca gaaaattttt acatgtggaa	6060
aaatgacatg gtggaaacaga tgcattgagga tataatcagt ttatggatc aaagcctaaa	6120
gccatgtgtaa atttaaccc cactctgtgt tagtttaag tgcactgatt tgaagaatga	6180
tactaataacc aatagtagta gcgggagaat gataatggag aaaggagaga taaaaactg	6240
ctctttcaat atcagcacaa gcataagagg taaggtgcag aaagaatatg catttttttta	6300
taaacttgtat ataataccaa tagataatga tactaccagc tataagttga caagttgtaa	6360
cacccatgcattc attacacagg cctgtccaaa ggtatcctt gagccatttc ccatacattt	6420
ttgtgccccg gctggtttg cgattctaaa atgtataat aagacgttca atggAACAGG	6480
accatgtaca aatgtcagca cagtacaatg tacacatgga attaggccag tagtatcaac	6540
tcaactgctgtt aataatggca gtcttagcaga agaagaggtt gtaatttagat ctgtcaattt	6600
cacggacaat gctaaaacca taatgtaca gctgaacaca tctgttagaa ttaattgtac	6660
aagaccccaac aacaatacaa gaaaaagaat ccgtatccag agaggaccag ggagagcatt	6720
tgttacaata gaaaaatag gaaatatgag acaagcacat tgtaacatta gtagagcaaa	6780
atggataaac actttaaaac agatagctag caaattaaga gaacaatttg gaaataataa	6840
aacaataatc tttaagcaat cctcaggagg ggacccagaa attgtaacgc acagttttaa	6900
ttgtggaggg gaattttct actgttaattc aacacaactg tttaatagta ctgggtttaa	6960
tagtacttgg agtactgaag ggtcaaataa cactgaagga agtgcaccaa tcaccctccc	7020
atgcagaataa aaacaaatataa taaacatgtg gcagaaagta gggaaagcaa tgtatgcccc	7080
tcccatcagt ggacaaatataa gatgttcattc aaatattaca gggctgctat taacaagaga	7140
tggtgtaat agcaacaatg agtccgagat cttcagaccc ggaggaggag atatgaggga	7200

caattggaga agtgaattat ataaatataa agtagtaaaa attgaaccat taggagtagc 7260
 acccaccaag gcaaagagaa gagtggtgca gagagaaaaa agagcagtgg gaataggagc 7320
 tttgttcctt gggttcttgg gagcagcagg aagcactatg ggccgcagcct caatgacgct 7380
 gacggtacag gccagacaat tattgtctgg tatagtgcag cagcagaaca atttgctgag 7440
 ggctatttag ggcgaacacgc atctgttgca actcacagtc tggggcatca agcagctcca 7500
 ggcaagaatc ctggctgtgg aaagataacct aaaggatcaa cagtccttgg ggatttgggg 7560
 ttgctcttggaa aaactcattt gcaccactgc tgtgccttgg aatgcttagtt ggagtaataa 7620
 atctcttggaa cagatttggaa atcacacgac ctggatggag tgggacagag aaatthaacaa 7680
 ttacacaagc ttaatacact ccttaatttga agaatcgcaa aaccagcaag aaaagaatga 7740
 acaagaatta ttggaatttag ataaatgggc aagtttggg aattggttta acataacaaa 7800
 ttggctgtgg tatataaaaat tattcataat gatagtagga ggcttggtag gtttaagaat 7860
 agttttgct gtactttcta tagtgaatag agttaggcag ggatattcac cattatcg 7920
 tcagacccac ctcccaaccc cgaggggacc cgacaggccc gaaggaatag aagaagaagg 7980
 tggagagaga gacagagaca gatccattcg attagtgaac ggatccttgg cacttatctg 8040
 ggacgatctg cggagcctgt gcctcttcag ctaccaccgc ttgagagact tactcttgat 8100
 tgtaacgagg attgttggaaac ttctggacg caggggggtgg gaagccctca aatattggg 8160
 gaatotccta cagtatttggaa gtcaggaact aaagaatagt gctgttagct tgctcaatgc 8220
 cacagccata gcagtagctg aggggacaga tagggttata gaagtagtac aaggagctt 8280
 tagagctatt cgccacatac ctagaagaat aagacagggc ttggaaagga ttttgctata 8340
 agatgggtgg caagtggtoa aaaagtagtg tgattggatg gcctactgtt agggaaagaa 8400
 tgagacgagc tgagccagca gcagataggg tggagcagc atctcgagac ctggaaaaac 8460
 atggagcaat cacaagtgc aatacagcag ctaccaatgc tgcttgcc tggctagaag 8520
 cacaagagga ggaggaggtg gttttccag tcacacctca ggtacctta agaccaatga 8580
 cttacaaggc agctgttagat cttagccact tttaaaaga aaagggggga ctggaaggc 8640
 taattcactc ccaaagaaga caagatatcc ttgatctgtg gatctaccac acacaaggct 8700
 acttccctga ttagcagaac tacacaccag ggcagggggt cagatatcca ctgacctttg 8760
 gatggtgcta caagcttagta ccagttgagc cagataagat agaagaggcc aataaaggag 8820
 agaacaccag cttgttacac cctgtgagcc tgcattggat ggtaccccg gagagagaag 8880
 tgtagatgtt gagggttgc acgcgcctag catttcattca cgtggcccgagact 8940
 cggagactt caagaactgc tgacatcgag ctgttacaa gggactttcc gctggggact 9000
 ttccaggag gctgtggcctg ggcggactg gggagtggcg agccctcaga tcctgcata 9060
 aagcagctgc ttttgcctg tactgggtct ctctggtag accagatctg agcctggag 9120

ctctctggct aactaggaa cccactgctt aagcctcaat aaagcttgcc ttgagtgctt 9180

c 9181

<210> 15

<211> 29

<212> RNA

<213> Homo sapiens

<400> 15

ggcagaucug agccugggag cucucugcc 29

<210> 16

<211> 52

<212> RNA

<213> Homo sapiens

<400> 16

uuuuuuuaggg aagaucuggc cuuccuacaa gggaggcga gggauuuuuc uu 52

<210> 17

<211> 9413

<212> DNA

<213> Homo sapiens

<400> 17

ttggggcga cactccacca tagatcactc ccgtgtgagg aactactgtc ttacacgaga 60

aagcgtctag ccatggcggtt agtatgagtg ttgtgcagcc tccaggaccc cccctccgg 120

gagagccata gtggtctgctg gaaccggta gtacaccgga attgccagga cgaccgggtc 180

ctttcttggta tcaacccgct caatgcctgg agatttgggc gtgccccgc gagactgcta 240

gcccggatgt gttgggtcgc gaaaggcctt gtggtaactgc ctgatagggt gcttgcgagt 300

gccccggag gtctcgtaga ccgtgcatca tgagcacaaa tcctaaacct caaagaaaaa 360

ccaaacgtaa caccaaccgc cgccccacagg acgttaagtt cccgggggtt ggtcagatcg 420

ttgggtggagt ttacctgttg ccgcgcaggc gccccaggtt ggggtgtgcgc gggactagga 480

agacttccga gcggtcgcaa cctcggtggaa ggcgacaacc tatccccaaag gctcgccggc 540

ccgagggttag gacctgggtt cagccccgggtt acccttggcc cctctatggc aacgagggtta 600

tggggtgtggc	aggatggctc	ctgtcacccc	gtggctctcg	gcctagttgg	ggcccccacag	660
accccccggcg	taggtcgctgt	aatttggta	aggtcatacg	tacccttaca	tgcggcttcg	720
ccgacacctat	gggttacatt	ccgcttgcg	gcgcggggct	agggggcgct	gccaggcccc	780
tggcacatgg	tgtccgggtt	ctggaggacg	gcgtgaacta	tgcaacaggg	aatctgccc	840
gttgctcttt	ctctatcttc	ctcttagctt	tgctgtcttg	tttggaccatc	ccagcttccg	900
cttacgaggt	gcgcaacgtg	tccggatat	accatgtcac	gaacgactgc	tccaaactcaa	960
gtatttgta	tgaggcagcg	gacatgatca	tgcacacccc	cgggtgcgtg	ccctgcgtcc	1020
gggagagtaa	tttctccgt	tgctggtag	cgctcactcc	cacgctcg	gccaggaaca	1080
gcagcatccc	caccacgaca	atacgacgccc	acgtcgattt	gctcggtggg	gcggctgc	1140
tctgttccgc	tatgtacgtt	ggggatctct	gccccatccgt	ttttctcg	tccagctgt	1200
tcaccttctc	acctcgccgg	tatgagacgg	tacaagattt	caattgctca	atctatcccg	1260
gccacgtatc	aggtcaccgc	atggcttggg	atatgatgt	gaactggtca	cctacaacgg	1320
ccctagtggt	atcgca	ctccggatcc	cacaagccgt	cgtggacatg	gtggcggggg	1380
cccactgggg	tgtcctagcg	ggccttgcct	actattccat	ggtggggaa	tgggctaagg	1440
tcttgattgt	gatgctactc	tttgcggcg	ttgacggca	caccacgtg	acagggggaa	1500
ggtaggcctc	cagcacccag	agcctcg	cctggctctc	acaaggccca	tctcagaaaa	1560
tccaaactcg	gaacaccaac	ggcagctggc	acatcaacag	gaccgctctg	aattgcaatg	1620
actccctcca	aactgggttc	attgctgc	tgttctacgc	acacaggttc	aacgcgtcc	1680
ggtgcccaga	gogcatggct	agctgccc	ccatcgatga	gttcgctcag	gggtgggg	1740
ccatcaactca	tgatatgcct	gagagctcg	accagaggcc	atattgtgg	cactacgc	1800
ctcgaccgtg	cgggatcg	cctgcgtc	agggtgtgg	tccagtgtat	tgcttactc	1860
cgagccctgt	tgtatgtggg	acgaccgatc	gttgcggc	tcctacgtat	agctgggggg	1920
agaatgagac	agacgtgt	ctacttagca	acacgcggcc	gcctcaaggc	aactggttt	1980
ggtagcgtg	gatgaacagc	actgggttca	ccaagacgtg	cggggccct	ccgtgcaaca	2040
tcgggggggt	cggcaacaac	accttggtct	gccccacgga	ttgcttccgg	aagcaccccg	2100
aggccactta	cacaaagtgt	ggctcgggc	cctgggtgac	acccaggtgc	atggttgact	2160
acccatacag	gctctggcac	tacccctgca	ctgttaactt	taccgtctt	aagtcagga	2220
tgtatgtggg	gggcgtggag	cacaggctca	atgctgc	caattggact	cgaggagagc	2280
gctgtactt	ggaggacagg	gataggc	aactcagccc	gctgctgt	tctacaacag	2340
agtggcagat	actgccc	tcc	ccttaccggc	cctgtccact	ggcttgc	2400
atcttcaccg	gaacatcg	gacgtgcaat	acctgtacgg	tatagggtcg	gcagttgt	2460
cctttgcaat	caaatgggag	tatatcctgt	tgctttcct	tcttgcggc	gacgcgc	2520

tctgtgcctg	cttgtggatg	atgctgctga	tagcccaggc	tgaggccacc	ttagagaacc	2580
tggtgtcct	caatgcggcg	tctgtggccg	gagcgatgg	ccttctctcc	ttcctcggt	2640
tcttcgtcgc	cgcctggta	atcaaaggca	ggctggtccc	tggggcggca	tatgctctct	2700
atggcgatg	gcccgttgc	ctgctttgc	tggcattacc	accacgagct	tatgccatgg	2760
accgagagat	ggctgcatcg	tgccggaggcg	cggttttgt	aggctggta	ctttgac	2820
tgtcaccata	ctataagggt	ttcctcgct	ggctcatatg	gtggttacaa	tatttatca	2880
ccagagccga	ggcgca	caagtgtggg	tcccccctct	caatgttcgg	ggaggcccg	2940
atgccccat	cctccttaca	tgccgggtcc	atccagagct	aatcttgac	atcacaaac	3000
tcctgctcgc	catactcggt	ccgctcatgg	tgctccaggc	tggcataact	agagtgcgt	3060
actttgtacg	cgctcagggg	ctcatccgt	catgcattt	agtgcggaa	gtcgctggag	3120
gccactatgt	ccaaatggcc	ttcatgaagc	tggccgcgt	gacaggtacg	tacgtatatg	3180
accatcttac	tccactgcgg	gattggccc	acgcgggcct	acgagacctt	gcggcggcag	3240
tagagccgt	cgtcttctct	gacatggaga	ctaaactcat	cacctggggg	gcagacaccg	3300
cggcgtgtgg	ggacatcatc	tcgggtctac	cagtctccgc	ccgaaggggg	aaggagatac	3360
ttcttaggacc	ggccgatagt	tttggagagc	aggggtggcg	gctccttgcg	cctatcacgg	3420
cctattccca	acaaacgcgg	ggcctgctt	gctgtatcat	cactagcctc	acaggtcggg	3480
acaagaacca	ggtcgatggg	gaggttcagg	tgctctccac	cgcaacgcaa	tcttcctgg	3540
cgacctgcgt	caatggcg	tgttggaccg	tctaccatgg	tgccggctcg	aagaccctgg	3600
ccggcccgaa	gggtccaatc	acccaaatgt	acaccaatgt	agaccaggac	ctcgccgt	3660
ggccggcgcc	ccccggggcg	cgctccatga	caccgtgcac	ctgcccgc	tcggacctt	3720
acttggtcac	gaggcatgt	gatgtcg	cggtgcgc	gcggggcgac	agcaggggg	3780
gcctgcttac	ccccaggccc	atctccat	tgaaggc	ctcgggtgg	ccactgctt	3840
gcccttcggg	gcacgttga	ggcatcttcc	ggcgtgtgt	gtgcacccgg	gggggttgcg	3900
aggcgggtgg	tttcatacc	gttgagtcta	tggaaactac	catgcgtct	ccggcttca	3960
cagacaactc	atccccctcc	gccgtaccgc	aaacattcca	agtggcacat	ttacacgctc	4020
ccactggcag	cgcaagagc	accaaagtgc	cggtgcata	tgca	gggtacaagg	4080
tgctcgctt	aaacccgtcc	gttgcgc	cattggc	tggagcgtat	atgtccaagg	4140
cacatggcat	cgagccta	atcagaactg	ggtaaggac	catcaccacg	ggcggcccc	4200
tcacgtactc	cacctattgc	aagtcc	ccgacgg	atgctccggg	ggccctatg	4260
acatcataat	atgtgatgaa	tgccactaa	ctgactcgac	taccat	ggcatacg	4320
ca	tcaggcag	acggctggag	cgccggctgt	cgtgc	accggccacgc	4380
ctccgggatc	gatcaccgt	ccacaccc	acatcgagga	agtggcc	c	4440

gagagattcc	cttctatggc	aaagccatcc	ccattgaggc	catcaagggg	ggaaggcatc	4500
tcatcttctg	ccattccaag	aagaagtgt	acagagctcg	cgcaaagctg	acaggcctcg	4560
gactcaatgc	tgttagcgtat	taccgggtc	tcgatgtgtc	cgtcataccg	actagcggag	4620
acgtcggtgt	cgtggcaaca	gacgctotaa	tgacgggttt	taccggcgac	tttgactcag	4680
tgatcgactg	caacacatgt	gtcaccaga	cagtcgattt	cagcttgat	cccaccttca	4740
ccattgagac	gacaacgctg	ccccaaagacg	cggtgtcg	tgcgcagcgg	cgaggttagga	4800
ctggcagggg	caggagtggc	atctacaggt	ttgtgactcc	aggagaacgg	ccctcaggca	4860
tgttcgactc	ctcggtcctg	tgtgagtgt	atgacgcagg	ctgcgcctgg	tatgagctca	4920
cgccccctga	gacctcggtt	aggttgcggg	cttacctaaa	tacaccaggg	ttgcccgtct	4980
gccaggacca	cctagagttc	tgggagagcg	tcttcacagg	cctcaccac	atagatgccc	5040
acttcttgc	ccagaccaaa	caggcaggag	acaacctccc	ctacctggta	gcataccaag	5100
ccacagtgt	cgccagggt	caggctccac	ctccatcg	ggaccaaatg	tggaagtgtc	5160
tcatacggct	aaagcccaca	ctgcatggc	caacgcccct	gctgtacagg	ctaggagccg	5220
ttcaaaatga	ggtcactctc	acacacccca	taaccaaata	catcatggc	tgcatgtcg	5280
ctgacctgga	ggtcgtaact	agcacctggg	tgcttagtagg	cggagtcctt	gcggctctgg	5340
cccgctactg	cctgacgaca	ggcagcgtgg	tcattgtggg	caggatcatc	ttgtccggga	5400
ggccagctgt	tattcccgac	agggaaatcc	tctaccagga	gttcgtatgag	atgaaagagt	5460
gtgcttcaca	cctcccttac	atcgagcaag	gaatgcagct	cggcagcaca	ttcaaacaga	5520
aggcgctcg	attgctgca	acagccacca	agcaagcgg	ggctgctgt	ccctgtgtgg	5580
agtccaagt	gcgagccctt	gaggtttct	gggcgaaaca	catgtggaa	ttcatcagcg	5640
ggatacagta	cttggcaggc	ctatccactc	tgcctggaaa	ccccgcgata	gcatcattga	5700
tggctttac	agcctctatc	accagccgc	tcaccaccca	aaataccctc	ctgtttaaaca	5760
tcttgggggg	atgggtggct	gcccaactcg	ctccccccag	cgctgcttcg	gtttcgtgg	5820
gcgcggcat	tgccgggtcg	gccgttggca	gcataggct	cggaaaggta	cttggacac	5880
ttctggcggg	ctatggggcg	gggggtggct	gcmcactcg	ggcccttaag	gtcatgagcg	5940
gcgagatgcc	ctccactgag	gatctggta	atttactccc	tgccatcctt	tctcctggcg	6000
ccctgggtgt	cggggtcg	tgcgacgca	tactgcgtcg	gcacgtggc	ccgggagagg	6060
gggctgtgca	gtggatgaac	cggctgatag	cggtcgatcc	gcggggtaac	cacgtctccc	6120
ccacgcacta	tgtgcccag	agcgacgccc	cggcgcgtgt	tactcagatc	ctctccagcc	6180
ttaccatcac	tcaagtgt	aaggatgtt	gggactggat	atgcacgg	tgctccacgc	6240
cttggccgg	ctcggtgcta	aaggatgtt	gggactggat	atgcacgg	ttgagtgtact	6300
tcaagacttg	gtccacttcc	aagctccctgc	cgcgggttacc	gggactccct	ttcctgtcat	6360

gccaacgcgg	gtacaaggga	gtctggcggg	gggatggcat	catgcaaacc	acctgcccatt	6420
gtggagcaca	gatcaccggaa	catgtcaaaa	atggctccat	gaggattgtt	ggccaaaaaa	6480
cctgcagcaa	cacgtggcat	ggaacattcc	ccatcaacgc	atacaccacg	ggccctcgca	6540
cgccctcccc	agcgccgaac	tattccaggg	cgctgtggcg	ggtggctgct	gaggagtacg	6600
tggaggttac	gcgggtgggg	gatttccact	acgtgacggg	catgaccact	gacaacgtga	6660
aatgccatg	ccaggttcca	gcccctgaat	tttcacgga	ggtggatgga	gtacggttgc	6720
acaggtatgc	tccagtgtgc	aaacctctcc	tacgagagga	ggtcgttattc	caggtcgggc	6780
tcaaccagta	cctggtcggg	tcacagctcc	catgtgagcc	cgaaccggat	gtggcagtgc	6840
tcacttccat	gctcaccgac	ccctctcata	ttacagcaga	gacggccaag	cgtaggctgg	6900
ccaggggtc	tccccctcc	ttggccagct	cttcagctag	ccagttgtct	gcgccttctt	6960
tgaaggcgac	atgtactacc	catcatgact	ccccggacgc	tgacctcatc	gaggccaacc	7020
tcctgtggcg	gcaggagatg	ggcgggaaca	tcacccgtgt	ggagtcagaa	aataaggtgg	7080
taatcctgga	ctcttcgat	ccgattcggg	cggtggagga	tgagagggaa	atatccgtcc	7140
cggcggagat	cctgcgaaaaa	cccaggaagt	tccccccagc	gttgc当地	tggcacgccc	7200
cggattacaa	ccctccactg	ctagacttcc	ggaaggaccc	ggactacgtc	ccccgggtgg	7260
tacacgggtg	ccctttgcca	tctaccaagg	cccccccaat	accacctcca	cgagggaaaga	7320
ggacggttgc	cctgacagag	tccaccgtgt	cttctgcctt	ggcggagctc	gctactaaga	7380
ccttggcag	ctccgggtcg	tcggccgttg	acagcggcac	ggcgactggc	cctcccgatc	7440
aggcctccga	cgtacggcgac	aaaggatccg	acgttgagtc	gtactcctcc	atgccccccc	7500
tcgagggaga	gccaggggac	cccgacctca	gacggggto	ttggtctacc	gtgagcgggg	7560
aagctggtga	ggacgtcgtc	tgctgctcaa	tgtcctatac	atggacaggt	gccttgcata	7620
cgcctatgcgc	tgcgaggag	agcaagttgc	ccatcaatcc	gttggcaac	tctttgtgc	7680
gtcaccacag	tatggtctac	tccacaacat	ctcgcagcgc	aagtctgcgg	cagaagaagg	7740
tcaccttga	cagactgcaa	gtcctggacg	accactaccg	ggacgtgctc	aaggagatga	7800
aggcgaaggc	gtccacagtt	aaggcttaggc	ttctatctat	agaggaggcc	tgcaaactga	7860
cgccttccaca	ttcggccaaa	tccaaatttg	gctacggggc	gaaggacgtc	cggagcctat	7920
ccagcagggc	cgtcaaccac	atccgcctcg	tgtggagga	tttgctggaa	gacactgaaa	7980
caccaattga	taccaccatc	atggcaaaaa	atgaggttt	ctgcgtccaa	ccagagaaaag	8040
gaggccgcaa	gccagctcg	tttatcgat	tcccagacct	gggggtacgt	gtatgcgaga	8100
agatggccct	ttacgacgtg	gtctccaccc	ttcctcaggc	cgtgatggc	ccctcatacg	8160
gattccagta	ctctcctggg	cagcgggtcg	agttcctggt	gaatacctgg	aaatcaaaga	8220
aatgccctat	ggcttctca	tatgacaccc	gttgcttga	ctcaacggtc	actgagaatg	8280

acatccgtac tggaaatca atttaccaat gttgtactt ggcccccggaa gccaggcagg 8340
 ccataaggc gtcacagag cggctttatg tcgggggtcc cctgactaat tcgaaggggc 8400
 agaactgcgg ttatcgccgg tgccgccaa gtggcgtgct gacgactagc tgccgcaaca 8460
 ccctcacatg ttacttgaag gccactgcgg cctgtcgagc tgcaaagctc caggactgca 8520
 cgatgctcgta gaacggagac gaccttgcg ttatctgtga gagtgcggga acccaggagg 8580
 atgcggcggc cctacgagcc ttcacggagg ctatgactag gtattccgccc ccccccgggg 8640
 acccgccccca accagaatac gacttgagc tgataacgac atgctccccc aatgtgtcg 8700
 tcgcgcacga tgcacccggc aaaagggtgt actacctcac ccgtgacccc accacccccc 8760
 tcgcacgggc tgcgtggag acagtttagac acactccagt caactccctgg ctaggcaata 8820
 tcacatgtatgc tgcgcacc ctagggcga ggtatgattct gatgactcat ttcttctcta 8880
 tccttcttagc tcaggagcaa cttgaaaaag ccctggatttgc tcaatctac gggccctgtt 8940
 actccattga gccacttgac ctacccatcaga tcattgaacg actccatggt cttagcgcac 9000
 tttcactcca cagttactct ccaggtgaga tcaatagggt ggcttcatgc ctcaggaaac 9060
 ttggggtacc gcctttgcga gtctggagac atcggccag aagtgtccgc gctaagctac 9120
 tgtcccaagggg ggggaggggct gccacttgac gcaagtaccc tttcaactgg gcagtaaaga 9180
 ccaagcttaa actcactcca atcccgctg cgtcccaacttgcg agacttgcgg ggctgggtcg 9240
 ttgctggta caacggggga gacatatatc acagccgttc tctgtccca ccccggttgg 9300
 tcacatgtatgc ctttctgtatggatggcat ctacccatcgc cccaaaccgg 9360
 gaaacggggag ctaaccactc caggccaaata ggccattccc tttttttttt ttc 9413

<210> 18

<211> 328

<212> RNA

<213> Homo sapiens

<400> 18
 uuggggcga cacuccacca uagaucacuc cccugugagg aacuacuguc uucacgcaga 60
 aagcgucuag ccauggcguu aguaugagug uugugcagcc uccaggaccc cccucucgg 120
 gagagccaua guggucugcg gaaccgguga guacaccggaa auugccagga cgaccggguc 180
 cuuucuugga ucaaccgcu caaugccugg agauuuugggc gugcccccgc gagacugcua 240
 gccgaguagu guugggucgc gaaaggccuu gugguacugc cugauagggu gcuugcgg 300
 gccccgggag gucucguaga ccgugcau 328

<210> 19

<211> 14

<212> RNA

<213> Homo sapiens

<400> 19

auuugggcgu gccc

14

<210> 20

<211> 27

<212> RNA

<213> Homo sapiens

<400> 20

gccgaguagu guugggucgc gaaaggc

27

<210> 21

<211> 340

<212> DNA

<213> Homo sapiens

<400> 21

atggcgag ggaagctcat cagtgggcc acgagctgag tgcgtcctgt cactccactc

60

ccatgtccct tgggaaggtc tgagactagg gccagaggcg gccctaacag ggctctccct

120

gagcttcagg gaggtgagtt cccagagaac gggctccgc gcgaggtcag actgggcagg

180

agatgccgtg gacccgccc ttcgggagg ggcccgccgg atgcctcctt tgccggagct

240

tggaacagac tcacggccag cgaagtgagt tcaatggctg aggtgaggtt cccgcagg

300

gacctcataa cccaaattcag accactctcc tccgcccatt

340

<210> 22

<211> 349

<212> DNA

<213> Homo sapiens

<400> 22

gaggaaagtc	cgggctcaca	cagtctgaga	tgattttagt	gttcgtgctt	gatgaaacaa	60
taaatcaagg	cattaatttgc	acggcaatga	aatatcctaa	gtctttcgat	atggatagag	120
taatttggaa	gtgccacagt	gacgttagctt	ttatagaaat	ataaaagggtg	gaacgcggta	180
aaccctcgat	gtgagcaatc	caaatttggt	aggagcactt	gtttaacgga	attcaacgta	240
taaacgagac	acacttcggt	aatgaagtgc	gtgttagacag	atggttatca	cctgagttacc	300
agtgtgacta	gtgcacgtga	tgagtacgt	ggaacagaac	gcggcttat		349

<210> 23

<211> 377

<212> DNA

<213> Homo sapiens

<400> 23						
gaagctgacc	agacagtcgc	cgcttcgtcg	tcgtcctctt	cgggggagac	gggcggaggg	60
gaggaaagtc	cgggctccat	aggcagggt	gccaggtaac	gcctgggggg	gaaacccacg	120
accagtgcaa	cagagagcaa	accggcgtat	gcccgcgca	gcgggatcag	gtagggtga	180
aagggtgcgg	taagagcgca	ccgcgcggct	gttaacagtc	cgtggcacgg	taaactccac	240
ccggagcaag	gccaaatagg	ggtcataag	gtacggcccg	tactgaaccc	ggtaggctg	300
ctttagccag	ttagcgattt	ctggcctaga	tgaatgactg	tccacgacag	aacccggctt	360
atcggtcagt	ttcacct					377

<210> 24

<211> 38110

<212> DNA

<213> Homo sapiens

<400> 24						
ccaccggta	cgatcttgcc	gaccatggcc	ccacaatagg	gccggggaga	ccggcgtca	60
gtggggcg	gcacggtcag	taacgtctgc	gcaacacggg	gttgactgac	ggcaatatc	120
ggctccatag	cgtcgccgc	ggatacagta	aaggagcatt	ctgtgacgga	aaagacgccc	180
gacgacgtct	tcaaacttgc	caaggacgag	aagggtcaat	atgtcgacgt	ccggtctgt	240
gacctgcctg	gcatcatgca	gcacttcacg	attccggctt	cggccttga	caagagcgtg	300
tttgacgacg	gcttggcctt	tgacggctcg	tcgattcgcg	ggttccagtc	gatccacgaa	360
tccgacatgt	tgcttcttcc	cgatcccgag	acggcgcgca	tcgaccgctt	ccgcgggccc	420

aagacgctga atatcaactt	480
ctttgtgcac gacccgttca	
ccctggagcc gtactccgc	
gaccccgca acatcgcccg	540
caaggccgag aactacctga	
tcagcaactgg catcgccgac	
accgcatact tcggcgccga	600
ggccgagttc tacatttcg	
attcggtgag cttcgactcg	
cgcgccaacg gtccttcta	660
cgagggtggac gccatctcg	
ggtggtgaa caccggcg	
gacccgagg cccgacggcag	720
tcccaaccgg ggctacaagg	
tccgcccacaa gggcggtat	
ttcccagtgg cccccaacga	780
ccaatacgtc gacctgcgc	
acaagatgt gaccaacctg	
atcaactccg gtttcatctt	840
ggagaaggc caccacgagg	
tggcagcgg cggacaggcc	
gagatcaact accagttcaa	900
ttcgctgctg cacgcccgg	
acgacatgca gttgtacaag	
tacatcatca agaacaccgc	960
ctggcagaac ggcaaaacgg	
tcacgttcat gcccaagcgg	
ctgttcggcg acaacgggtc	1020
cggcatgcac tgtcatcgt	
cgctgtggaa ggacggggcc	
ccgctgatgt acgacgagac	1080
gggttatgcc ggtctgtcg	
acacggcccg tcattacatc	
ggcggcctgt tacaccacgc	1140
gccgtcgctg ctggccttca	
ccaacccgac ggtgaactcc	
tacaagcggc tggttcccg	1200
ttacgaggcc ccgatcaacc	
tggtctatag ccagcgcac	
cggtcggcat gcgtgcgc	1260
cccgatcacc ggcagcaacc	
cgaaggccaa gcgctggag	
ttccgaagcc ccgactcg	1320
ggcaacccg	
tatctggcgt tctcgccat	
gctgatggca	
ggcctggacg gtatcaagaa	1380
caagatcgag ccgcaggcgc	
ccgtcgacaa ggatctctac	
gagctgcccgc	1440
cggaagaggc cgcgagtatc	
ccgcagactc cgaccagct	
gtcagatgt	
atcgaccgtc tcgaggccga	1500
ccacgaatac ctcaccgaag	
gaggggtgtt cacaacac	
ctgatcgaga cgtggatcag	1560
tttcaagcgc	
gaaaacgaga tcgagccgt	
caacatccgg	
cgcattccct acgaattcgc	1620
gctgtactac gacgtttaag	
gactctcgc agtccgggt	
tagagggagc ggcgtgtcg	1680
tgccagggcg ggcgtcgagg	
ttttcgatg ggtgacgg	
gcccggcaacg ggcgcgc	1740
caccgtcg	
aagagcccgt ttaagaacgt	
tcaaggacgt	
ttcagccgg	1800
tgccacaacc cgcttggca	
tcatctcccg accgcccagc	
gggttgtctt	
tcacatcg	1860
cgaaactcaa gccacgtcg	
cgcccaggcg tgcgtcg	
gccggttcag	
gttaagtgtc gggattcgt	1920
cgtgcggcg	
ggcgtccacg ctgaccaacg	
ggcagtc	
ctcccgaaaca ctttgcgcac	1980
taccgcctt gcccggcg	
tcacccgtag gtatgtcc	
aggaattccc caccgtcg	2040
gtttcgccag ccggccgcga	
ccgcgaccgc attgagctgg	
cgcgggg	2100
ccggcagctg gtcgggg	
ttgcccgc	
ccaacaccag cgcg	
gcccgggtgg cggtcagcca	2160
ggcctgacgg	
agcagtc	
cgatcgactc	
caggattgc agcgtcgagg	2220
tgttgtgc	
ggcgggaaacc	
tggtgcgc	2280
catgttagctg cagcaactgc	
acggtccatt cgatgtcg	
cagtccggc	
cggccagtt tggtgtgtgt	2340
gttgggtcg	
gcaccgcg	
gcaaccgc	
ggactcgata	

cgggccttga	tgcggcgaat	ctcgccgacc	gagtcagcgg	acacaccgtc	gggcggatac	2400
cgcgtttgt	cgaccatccg	taggaatcgc	tgacccaact	cggcatcgcc	ggcaaccgcg	2460
tgtgcgcgta	gcagggcctg	gatctccat	ggctgtgccc	actgctcgta	gtatgcggcg	2520
taggacccc	gggtgcggac	cagcggaccg	ttgcggccct	cgggtcgcaa	attggcgtcg	2580
agctccagcg	goggatcgac	gctgggtgtc	cccagcagcg	cccgaaacccg	ctcgccgatc	2640
gatgtcgacc	atttcaccgc	ccgtgcac	tcgacgcccgg	tggccggctc	acagacgaac	2700
atcacgtcgg	catccgaccc	gtagccaaac	tcggcaccac	ccagccgacc	catgccatg	2760
accgcgatgg	ccgcgggggc	gcgatcg	tcgggaaggc	tggccggat	catgacgtcc	2820
agcgccgcct	gcagcaccgc	cacccacacc	gacgtcaacg	cccgacacac	ctcggtgacc	2880
tcgagcaggc	cgagcagg	cgccgaaccg	atgcggcca	gctctcgacg	acgcagcgtg	2940
cgcgccgcgg	cgatggcccg	ctccgggtcg	ggtagcggc	tcgcccggc	gatcagcgcc	3000
cgagccacgg	cgccgggctc	ggtctcgagc	agttcggc	ccgcaggccc	gtcctcgat	3060
tgctggatga	cccgccgcgc	gcbcataac	agatccggca	catacggca	gttaccaag	3120
acatgcatga	gccgcttggc	caccgcgggc	ttgtccgc	gcgtggccag	gtaccagtt	3180
tcgggtggcca	gcccctca	gagccgcgg	taggcagca	gtccgcgtc	ggatcgggg	3240
gcatacgaca	tccagtccag	cagcctggc	agcagcaccg	actgcacccg	tccgcgcgg	3300
ccgctttgat	tgaccaacgc	cgacatgtgt	ttcaacgcgg	tctgcggtcc	ctcgtagccc	3360
agcgccggca	gccggcgccc	cgcggctcc	aacgtcatgc	cgtggcgat	ctccaacccg	3420
gtcggccga	tcgattccag	cagcgggtga	tagaagagtt	ttgtgtgtaa	cttcgacacc	3480
cgcacgttct	gttcttgag	ttcctccgc	agcaccccg	ccgcatacg	tcggccatcg	3540
ggccggatgt	ggccgcgcg	ccgcagccag	cgcactgcct	cctcgcttc	ggatcggga	3600
agcaggtggg	tgcgcttgag	ccgctgcaac	tgcaatcggt	gctcgagcag	cctgaggaac	3660
tcatacgacg	cggtcatgtt	cgccgcgtcc	tcacgcccga	tgtagccgc	ttcgcccaac	3720
gccgccaatg	cgtccaccgt	ggacgcccacc	cgtaacgact	cgtcgctacg	ggcatgaacc	3780
agctgcagta	gctgtacggc	gaactccacg	tcgcgcatac	cgccgctgcc	gagtttgagc	3840
tcgcggccgc	ggacatcg	gggcaccagc	tgctccaccc	gccgccc	catgcacc	3900
tcgaccacaa	agtcttcgc	ctcgcaagg	cgccacacca	tcggcatcaa	ggcggtcagg	3960
taacgctcgc	caagttccgc	gtcgccaaac	actggccgt	ctttcagca	cgcctgaaac	4020
tcccgagg	tggccagcg	ctggtagtag	gcgtgtgc	actcgagcgt	acggaccagc	4080
tcccggttgc	gccccctccgg	acgcaggcg	gcgtccac	cggaaaaggc	cgccgaggcc	4140
acccgcac	tctcgctggc	cacgcgcgc	ttgcgcgggt	cggagcgtc	ggcaacgaat	4200
atgacatcg	cgtcgctgac	gtatcgt	tcgcgcgcac	cgcacttgcc	catcgat	4260

accgcaggc	gccccggcgg	gtgctcgccg	cacacgctcg	cctcgccac	gccccggcgg	4320
gccgcaggag	cggcgcccg	ggcgcccgcc	aggcggtcg	ccaccacgg	aatggcagc	4380
accgggtcg	cctcgaccgt	cgcggccagg	tcgagagcg	ccagcattag	cacgtatcg	4440
cggtaactgg	ttcgcaatcg	gtgcacgagc	gagccggca	taccctccga	ttccctcgac	4500
cactcgacga	acgaccgctg	cagctggta	tgggacggca	gtgtgacctt	gccccggcagc	4560
aatttccagg	actgcggatg	ggcgaccagg	tgcgcggcc	acgcccagcga	cgagcccgac	4620
accgagaaca	gccccggcgcg	cagactgcgt	tcgcccggca	gagccggcgtt	gagctcggtcc	4680
catccggtgt	ctggattctc	cgacagccgg	atcaaggcgc	gcagcggcggc	atccggcgtcc	4740
ggagcgcgtg	acagcggatc	cagcaggctcg	acgtgcgcct	gatccctcg	ccgatcccac	4800
cccagctgag	ccagacgctc	accagcaggg	gggtcaacta	atccgagccg	gccaacgctg	4860
ggcaacttcg	gcccgtgcgt	ggcgagttt	gtcacggatc	cgacggtagc	gcaaaggcgcg	4920
tcggcgtcgg	atcaaccgg	agatctggc	tacagcgcaca	ggttaggtcg	cagctcgat	4980
ggcgtgacgt	ggctcggt	gttcgcccac	tccgtcgct	tgttgcgca	aaaaaagtca	5040
aaaacgtgct	cccccaaggc	ctccgcgcac	agttcggagg	cctccatggc	gccccggcga	5100
ctatccaaac	tggacggcaa	ttctcggtac	cccatcgctc	ggcggtcctc	gggtgtgagg	5160
tcccatacgt	tgtcctcgcc	ctgcgggccc	agcacgtaac	ccttctctac	accccgcaat	5220
cccgccggca	gcagcacggc	aatgtcaga	tagggattgc	acgcccata	agggtcggt	5280
acttcgaccc	gcccgcacga	ggtcttgc	ggcgtgtaca	tcggcaccc	cactagggcg	5340
gatcggttgg	cggccccc	cgacgcggcc	gtggcgctt	cggccccc	caccagccgc	5400
ttgttaagagt	tgacccactg	atttgtgacc	gacgtatct	cgcaacgcgt	ctccaggatc	5460
ccggcgtat	acgatttacc	cacttccgac	agctgcagcg	gatcatcagc	gctgtggaa	5520
gcgttgcacat	caccctcgaa	caggctcatg	tgggtgtca	tcggcagcc	cggtgtcg	5580
ccgaatggct	tgggcatgaa	cgacgcccgg	gccccctt	ccagcgcac	ttctttgat	5640
acgtacgg	aggcatcac	gttgcagcc	atcgacagag	cgtcggcaaa	ccgcaggatcg	5700
atctcctgct	ggccgggtgc	gccttcgt	tggctgact	ccaccgagat	gccccatgaa	5760
tccagggcat	cgatcggt	gccccggaa	ttcaaggccg	agtcgtgcac	cgcttggat	5820
aaatagccgg	cgttgcac	cgggacgggc	accgaccgt	cctcggtcc	gggttgcgc	5880
aggaagaact	cgatccgg	atgcacgt	caggagaagc	cgatccgc	ggccctcg	5940
agctgcgc	gcaacacgt	ccgcgggtcc	gccccacgac	gccccggc	cgccatgg	6000
atgtcgaaa	acatccgcgc	ttagtgg	tggccggaa	tggggccca	gggcagcacc	6060
tggaaagg	acgggtccgg	gtgcggcacc	gtatcgatt	ccgagaccc	cgcaaaagccc	6120
tcgatcgagg	atccgtcgaa	gccgatgcct	tcctcgaa	cgccctcg	ttcggtgg	6180

gcgatggcga ccgactttag	6240
gaaaccgagc acgtctgtga	
accacagccg gacgaagcgg	
atgtcgctt ctccagggt	6300
acgaagaacg aattccttct	
gtcggtccat acctcgaa	
ctatgcaactg tctgttaaaa	6360
ccgtgttacc gatgcccggc	
cagaagcggt gcggggcggc	
ccgcaagggg agtgcgcgg	6420
gagttcaggg cgccgaccgc	
agactcgctg gcggcaaggt	
cccgtcgaga aaatagtgc	6480
tcaccccgaga gtccacacac	
tggttgccat cgaacaccgc	
agtgtgttgg gtgccgtcg	6540
aggtgatcag cggtgccccc	
agctggcggg ccaggtctac	
cccgactga tacggagtgg	6600
ccgggtcgtg ggtggtgac	
accacgacga ccttgc	
cccgccggc gccgcgggt	6660
gcccgcgtcg	
cgttgccggc accggccaca	
gcgcgcacag atcgcg	
ggggatccgg tgaactgccc	6720
gtagctaagg aacggggcga	
cctgacggat	
ccgttggtcg gcccgcaccc	6780
aggccgctgg atcggccgg	
gtgggcgcat cgacgcac	
gaccgcgttg aacgcgtcct	6840
ggtcgttgct gtagtgc	
tctgcataccc gcccgtata	
gtcgtcggca agcaccagca	6900
agtcgcggc gtcgtgc	
cgctgcagcc ccagcagacc	
actggtcagg tacttccagc	6960
gctgagggtc gtacagcgc	
tttagtggtc ccgtcg	
gtcggcgtag ctcaggccac	7020
gtggatccga cgttta	
ccgcgtca ccagcgg	
ggcttctgca ccagcgg	
aaccaggcg	7080
tgtagcggt tgaccactg	
ggccgagtc gtgcccagag	
ggcaggccgg	
cgagcggcg	7140
cagtcggcg	
cgtagtcatt gaaagcg	
tgaaatccc ccattgg	
gatgcttcc tcgattggc	7200
taacggctgg atcgatagc	
ccgtcgagga ccatcg	
cacatgagta ccgaaccgtt	7260
ccaggtaa	
ggtgcccac tcggtgcc	
agctgtatcc	
gaggtatgg atctgatcg	7320
cacctaacgc ttggcga	
accatgt cccgtgc	
ggacgcggta ccgatattgg	7380
ccaagaagct gaagccc	
atcgtc acgcgtcg	
caactgccc tagacctgtt	7440
cgacgtgg	
gacacggcc ggactgt	
cgccatcg	
atcgccgg tacgcgtcg	7500
actcgccgt ggtgc	
gacacgc	
gggtcgagtg	
gccgaccct ctcgggtcg	7560
agcccaccag gtcga	
ggagaatgt cgggtcg	
gatcgccggt gccatagcg	7620
cgaccatgt	
gaccgac gaccgt	
gaccatgtc gaccg	
gccccggc ccccagg	
gaccagcgt gctccga	7680
atcgtccgt	
cgccggacg cggatcac	
ccaacttcgc	
ttgtgtccca cccgg	7740
ttgtgtccca cccgg	
cgtagtcgac gggacgg	
accgtcg	
agcgtgc	
gcaatttcg ctggtg	7800
cgatgaactc gggcag	
cttccaaactt gttcg	
gttgcggc	
cacgaccggc	7860
gcacccggg	
tttggccggc	
gcccgggt	
tcagtcgc	
cgcccaacgg	
gggcgtgct	7920
agggcagtc	
cgccgagc	
caacccga	
gacagcagc	
ccgagctaa	
cggtctcg	7980
cgccacatgg	
ccgcattcg	
ctcaccggcg	
aatacgt	
acggcgc	
atgatcacac	8040
cttcgttct	
tcgccccgt	
agcacttggc	
gccgctggc	
ggcgtgg	
cgccgattaa	8100
atacgcgc	
acgtactcg	
caatgcag	
gtcgccctgg	
aataccaccg	

tgtgctgggt tccgtcgaag gtcagcaacg aaccgcgaag ctggttcgcc aggtcgaccc	8160
cggccttgt a cggcgtcgcc gggcatggg tggtgatac caccaccgtc ggcactaggc	8220
cgggcgcga gacggcatgg ggctgacttg tggtgccac cggccagaac ggcaggtgc	8280
ccagcggcgc atcaccggtg aacttccgt agctcatgaa cggtgcgatc tcccgccgc	8340
ggcggcttc gtcgatgacc ttgtcgatcg cggtaaccgg gggctgatcg acgcaattga	8400
tcgcccacccg cgctcaccg gaattttgt agcggccgtg cgagtcccg a cgcattaca	8460
tgtcgccag agccagcagg gtgtctccgc gattgtcgac cagctccgac agccgtcgg	8520
tcaagtgtt ccacagattc ggtgagtaca gcgcataat ggtgccacg atggcgtcgc	8580
tataactcag cccgcgcgga tccttcgtgc gcgcggcct gctgatcctc gggttgtccg	8640
ggtcgaccaa cggatcgacc aggctgttgt agacctcgac ggctttggcc gggcggcgc	8700
ccagcggca gcccgcgttc ttggcgcagt cggcggcata gttgttgaac gcgtcctgga	8760
agcccttggc ctggcgcage tccgcctcga tggatcgac attgggtcg acggcaccgt	8820
cgagaatcat tgccgcacc cgctgcggaa attcctcgac atacgcggag ccgatccggg	8880
tgccgtacga gtagcccagg taggtcagct tgcgtcgcc caacgcgcg cgaatggcat	8940
ccaggtcctt ggcgacgttg accgtcccgaa catggccag aaagtttttgc cccatcttgc	9000
ccacacagcg accgacgaat tgcttggctc cgttctcgat gtgcgcacca ccctccggc	9060
tgttagtcaac ctgcggctcg gcccgcagcc ggtcggtgc ggcacatggag ttgcaccaga	9120
tcgcccggcg ggacgacgccc accccgcggg ggtcgaaccc aaccaggcgt aacctttcg	9180
gcacccgctt cggcaatgtc tggaaagacgc ccaaggcggc ctcgataaccg gattcggcgg	9240
gtccaccggg atttatgacc agcgaaccga tcttgcgtcc cgtcgccgaa aacgcaatca	9300
gcccacgcg cgcacgtca ccatcgccggc ggtcgtagtc gaccggtaca gcgagcttgc	9360
cgcataacgc gcccgcgggg atctttactt gcccgttgc cgcaccgcac ggtgtccact	9420
ccaccggctg gcccagcttc ggctccgcca tacgagcgcg tccccgacc acgcggatgc	9480
agcccacaag aaccaacgccc acggcggcga ggcggccca gatcaacacgc atgcgcgcga	9540
tcttgcgtcg gcgagacagc ctcatgccc caatgctgcc agagcagacc cgagatctgc	9600
gccagcggcc accgtcgcc gactaaccgg ccgctgcccag cagtcctgccc atgcggatgc	9660
gcgaactcgt cggccatccc ccatacgtcc ggtAACAGAT ccggcaaga caccgaccccg	9720
tcgaccggat cccgtccacgg cgcgtcgcc tccggcgtgc acaactgcga catcagggttgc	9780
gcccgtccac cccgtccacgg cccgtcatggt gcacccgttgc catcgccgc gggcgatccc	9840
cgcgtccac cccgtccacgg acgaacccat ctcccacggc ggtcgccggc agcgacgcga	9900
tgtggccgca gatctccgag agttcgccccc gcccggccgg cgcggcaac ccgatgcccgt	9960
gcaagtgcacg atcgatgtga ggtcaaggt tcagcgcact gctggcaacg ttttccgaa	10020

accgcggcct cgcccttgate tggagtctaga acgcgtcactc cagccggta aaggcgtaac 10080
 ccatgctcga gcaaacatgc atgggctgag tggacgttc cagacacagc aactggcg 10140
 caggccactg agccgctgca tgcgcgtatgg tatgcgcgtatgg ggggccccgg gcgctctga 10200
 ggggaagaag tggcagactg tcagggtccg acgaacccgg ggaccctaac gggccacgag 10260
 gatcgacccg accaccatta gggacagtga tgtctgagca gactatctat ggggccaata 10320
 ccccccggagg ctccggggccg cggaaccaaga tccgcacccca ccacctacag agatggaagg 10380
 ccgacggcca caagtggcc atgctgacgg cctacgacta ttgcacggcc cgatcttcg 10440
 acgaggccgg catcccggtg ctgctggcgtc gtgattcggc ggccaacgtc gtgtacggct 10500
 acgacaccac cgtgcccgtc tccatcgacg agctgatccc gctggtccgt ggcgtggc 10560
 ggggtgcccc gcacgcactg gtcgtcgccg acctgcccgtt cgccagctac gaggcgggc 10620
 ccaccgcccgc gttggccgccc gccaccgggt tcctcaagga cggcggcgca catgcggc 10680
 agctcgaggg cggtgagcgg gtggccgagc aaatcgctg tctgaccgacg gcccgcattcc 10740
 cggtgatggc acacatcgcc ttccacccgc aaagcgtcaa caccttggc ggcttccggg 10800
 tgcagggccg cggcgacgccc gccgaacaaa ccatcgccga cgatcgcc gtcgcccgaag 10860
 cccggcggtt tgccgtcgatg atggagatgg tgcccgccga gttggccacc cagatcaccg 10920
 gcaagcttac cattccgacg gtcgggatcg gctggccgtt caactcgacg gcccaggatcc 10980
 tggtatggca ggacatggcc gggttcagcg ggcaccaagac cggccgttgc gtcaaacgg 11040
 atgcccgtt cggtggtgaa ctacgcccgtg ctgcaatgca atacgcccggg gagggtggccg 11100
 gcccgggtt ccccgctgac gaacacagtt tctgaccaag cccgatcgccg 11160
 ggcattgcgg tggcgccctg gatgcccgtg acgcccggatt gcccggcgacg acggccacgc 11220
 gggaccatc ggcgtcgctg tcgcccgtt gggccgggtt gagccacagac attcgatgt 11280
 cccaaacacca tccgcccacag cccaaattgtat gtggcactct atgcattgcct atccccgacc 11340
 aaccaccacc gcggcgacgc atcatgaccg gaggcgaaga tgccagtata ggcgcggcaga 11400
 ccagcgccgc atctggaggt cgagcgacg ttcgacgtga tcgactcgac ggtgtcgccg 11460
 tcgttcgagg gcatcgccgc ggtgggtcg gtcgagcgtc cgccgacccca gcaatcgac 11520
 gcccgggtact tcgacacacacc gtcgcacgac ctggcgccga accagatcac ctgcggcg 11580
 cgcacccggcg ggcgcggacgc cggctggcat ctgaagctgc cggccggacc cacaacgc 11640
 accgagatgc gaggcaccgct gtccgcacatca ggcgcacgctg tgccggccga gttgtggat 11700
 gtggtgctgg cgatcgccgtc cgaccacggc gttcagccgg tcgcccggat cagcactcac 11760
 cgcgaaagcc agatcctgtca cggcgccggg ggcgcacgc tggcggaaatt ctgcacacgac 11820
 gacgtcaccg catggtcggc cggggcattc cacgcgcgtg gtgcacgcga caacggccct 11880
 gcccggacgc agtggcgacgc atgggaactg gaactggcata ccacggatgg gaccggcgat 11940

accaagctac tggaccggct agccaaccgg ctgctcgatg ccggtgccgc acctgcccgc 12000
cacggctcca aactggcgcg ggtgctcggt gcgacctctc ccggtgagct gccaacggc 12060
ccgcagccgc cggcggatcc agtacaccgc gcggtgtccg agcaagtcga gcagctgctg 12120
ctgtgggatc gggccgtgcg ggccgacgccc tatgacgccc tgcaccagat gcgagtgacg 12180
acccgcaaga tccgcagctt gctgacggat tcccaggagt cgtttggcct gaaggaaagt 12240
gcgtgggtca tcgatgaact gcgtgagctg gccgatgtcc tggcgttagc ccgggacgccc 12300
gaggtaactcg gtgaccgcta ccagcgcgaa ctggacgcgc tggcgcggaa gctggtaacgc 12360
ggccgggtgc gcgagcgcct ggtagacggg gcgccggcgc gataccagac cggcgtgcgg 12420
cgatcaactga tcgcattgcg gtcgcagcgg tacttccgtc tgctcgacgc tctagacgcg 12480
cttgtgtccg aacgcgcccga tgccacttct ggggaggaat cggcaccggat aaccatcgat 12540
gcggcctacc ggcgagtcgg caaagccgca aaagccgcaa agaccgcgg cgaccaggcg 12600
ggcgcaccacc accgcgcacga ggcattgcac ctgatccgca agcgcgcgaa gcgattacgc 12660
tacaccgcgg cggctactgg ggcggacaat gtgtcacaag aagccaaggat catccagacg 12720
ttgctaggcg atcatcaaga cagcgtggtc agccggaaac atctgatcca gcaggccata 12780
gccgcgaaca ccgcggcgcg ggacaccttc acctacggtc tgctctacca acaggaagcc 12840
gacttggccg agcgctgcgc ggagcagctt gaagccgcgc tgcgcaaact cgacaaggcg 12900
gtccgc当地 caagggattt agccgc当地 ggcggacga gttggctgt aagccggatt 12960
ctgttccgc当地 ccgc当地 acgc当地 caagctaactc gccgc当地 ccgc当地 acgc当地 13020
caccgttacc gggcgc当地 acgc当地 ccgc当地 acgc当地 ccctcaactc 13080
cctgc当地 ccgc当地 acgc当地 ccgc当地 acgc当地 ccctcaactc 13140
gccacccccc当地 tcacccggaa tgctggcgc当地 ctcttaccgc当地 accgttccac ccttgc当地 13200
acgaggatgg cggctgttt tctgtggcact tttccgc当地 gtcacccggg attgtccgtta 13260
gcaatcaccc tgctctgtga agtcoggact ttccctcgact cgacgctgaa cctcgtgaat 13320
ccacacaaggc cctacgc当地 ccgc当地 ccgc当地 acgc当地 catccgc当地 gaccacgc当地 13380
ccccgctggg cggcgtgc当地 gcaactgtga ccgc当地 acacggctag tcggacagcc 13440
gatccggccgg gcaactgc当地 tcgtggactg gtgacacggat gggacaaacgc cgtc当地 acgc当地 13500
ggcgc当地 acgc当地 acgc当地 tgccgaggc当地 agcgactacg gtggc当地 acgc当地 gctacccgg 13560
ctgatcaccc ccggc当地 gaggc当地 cggccggctg cgc当地 acgc当地 acgc当地 cggc当地 ctgttt 13620
cgctcgactggg tcgatatggc atccaaacgc当地 tacggc当地 ccgc当地 acgc当地 atatccat 13680
gccccctatc ccgactgtatc gagcgtctca acgc当地 acgc当地 gtatccaaa ctgctgc当地 13740
tagcgc当地 cttggc当地 aaactggc当地 gggaggc当地 ctggccagac acgc当地 ttgtatc 13800
actggc当地 gagctgtatc gccc当地 cggccggcc当地 aaacccgatc acacgc当地 ctgttaaggt 13860

acggcaccaa cgactggaac gccctacacc aggatctcta cggcgagttg gtgttccgc 13920
 tgcaggtgg tgcacacccg agcgatccgg aaaccgacta caccggcggc gagttcctgc 13980
 ttgtcgaaca gcggcctcgc gcccaatccc ggggtaccgc aatgcaactt ccgcaggac 14040
 atggttatgt gttcacgacc cgtgatccggc cggtgccggac tagccgtggc tggtcggcat 14100
 ctccagtgcg ccatgggctt tcgactattc gttccggcga acgctatgcc atggggctga 14160
 tctttcacga cgcaagctga ttgcacgcca tctatagata gcctgtctga ttcaccaatc 14220
 gcacccgacga tgccccatcg gcgtagaact cggcgatgtc cagcgatgcc agatcaagat 14280
 gcaaccgata taggacgccc gacccggcat ccaacgcccag ccgcaacaac attttgc 14340
 gcgtgacatg tgacaccacc agcaccgtcg cgcctcgta gccaacgatg atccgatcac 14400
 gtcccccccg aaccggccgc agcacgtcgt cgaagcttc cccacccggg ggcgtgatgc 14460
 tggtgtcctg cagccagcga cggtgcaact cgggatcgcc ttctgcccgc tccgcgaacg 14520
 tcagccctc ccaggcgccg aagtccgtct cgaccaggcgtc gtcatcgacg accacgtcca 14580
 gggccaggcgc tctggcggcg gtcaccgcgg tgtcgtaagc ccgctgttagc ggcgaggaga 14640
 ccaccgcagc gatccggccg cgccgcgcca gataccggc cgccgcacca acctggcgcc 14700
 accccaccc tttcaacccc ggggtccgc gccccata gccgcgttgc tccgacagct 14760
 ccgtctgccc gtggcgcaac aaaagtatgc ggggtgggtgt acccgccggc cccgtccagc 14820
 cgggagatgt cggtgactcg gtcacacgaa ttttggcagg atccgcatacc gccgcagccg 14880
 attgcgcggc ggcgtccatc gcgtcattgg ccaaccggcgtc tgcatacgtg ttccggcac 14940
 gcggaaacca ctcgtatgg atcctgcgaa actggacgc caacgcctga gcctggacat 15000
 agagcttcag cagatccggg tgcttgaccc tccaccgcggc ggacatctgc tccaccacca 15060
 gcttggagtc catcagcacc gcggcctcgg tggcacctag tttcacggcg tcgtccaaac 15120
 cggctatcag gcccggatcg tccgcacgt tggtcgccgc ccggccgatc gcctgcttgg 15180
 actcggccag cacggtgag tgatcgccgg tccacaccac cgcgcgtat ccggccggc 15240
 cgggattgcc cccgcgtccg cgtccggctt cgatgacaac tttcactcct caaatccctc 15300
 gagccgcaac aagatcgctc cgcattccgg gcagcgcacc acttcacccct cggccggccgc 15360
 cgagatctgg gccagctcg cgcggccgat ctgcgtccgg caggcaccac atcgatgacc 15420
 ttgcaaccgc ccggccctcg gcccgcctcc ggcccgtgt ctttcgtaga gccccgcaag 15480
 ctcggatca agtgtcgccg tcagcatgtc gcgttgcgt gaatgttggt gccggcttgc 15540
 gtcgattcg gcaagtgcct cgtccaaagc ctgcgtccgg cggccggatcg cggccgc 15600
 cgcttggagc gcccggact cggccgtctg ttgagcctgc agtcctcgcc ggcgttccag 15660
 cacctccagc agggcatctt ccaaactggc ttgacggcgt tgcaagctgt cgagctcg 15720
 ctgcagatca gccaattgtc tggcgtccgt tgcacccgaa gtgagcaacg accggcccc 15780

gtcgccacgc ttacgcacccg catcgatctc cgactcaaaa cgcgacacacct ggccgtccaa 15840
gtcctccgcc gcgattcgca gggccgccc cctgtcggttgc gggcggtgt gtcggccctg 15900
cacctgctgg taagccgccc gctgcggcag atgggtagcc cgatgcgcga tccgggtcag 15960
ctcagcatcc agttcgcca attccagtag cgaccgttgc tgtccactc cggctttcat 16020
gcctgatctc tcccagtttc gtgatcgagg ttccacgggt cggtgacatc ggtgcacaca 16080
cgcacccggca gcgacgcgccc gaaatgagac cgcaacactt cggcggcctg gccgcaccac 16140
ggaaattcgc ttgcccatacg cgacgtcg atcaggggca cttgcgaagc tcggcaatgc 16200
tcgtcggctg gatgatgtcg cagatcgcc gtaacgtacg cttgcacgtc cggcggcggcc 16260
acggtgccaa gcaacgagtc cccggcgccg cccgaccccg cgacccgcga caccagcagg 16320
tcgggatccc cggcggcgccg cacaccggcgc gcaacgcggc ctccagacgg 16380
gcaacaaagg tgcgcagcgg ttccgggttt ggcagtctgc caatccggcc taacccgctg 16440
ccgacccggcg gtggtaccag cgacgtatcg tcgaatgcgg gtcctcgta agggtgcgcg 16500
gcgcgcacatcg cccgcaacac ctcggcgccg gtcgtgcgg gtgcacgc acgcacccgg 16560
tcctcgccca cccgttcgac ggtaccgacg ctgcctatgg cgggcacgc cccgtcgatcg 16620
gccaggaact gcccggtacc cgacgtaccc cagctgcagt gcgagtagtc gcgcataatgg 16680
ccggcaccgg cctcaaagac cgctgcccgc accgcctctg agttctcgcg cggcacatacg 16740
atgacccact tgcgagatc ggccgctccg ggcaccgggt cgagaacggc gtcgacggtc 16800
agaccaacag cgtgtgccag cgacgtcgac acacccggcg acgcccggatc ggcgttggtg 16860
tgcgoggtaa acaacgagcg accggtccgg atcaggcggt gcaccacgc accctttggc 16920
gtgttggccg cgaccgtatc gacccacgc agtaacaacg ggtggcgcac caatagcagt 16980
ccggcctggg gAACCTGGTC caccaccgc ggcgtcgctt ccaccgcac ggtcaccgaa 17040
tccaccacgt cgtcggttc gcccacacc agacccacccg aatcccacga ctggcaagc 17100
cgccggcggtt aggcctggtc cagcacgtcg atgacatcg ccagccgcac actcatcgcc 17160
gtcctccacg ctttgcac tcggcgatcg ccggcaccag cacggccac tccggcgca 17220
ccggcggcccg caggtaccgc ggcgtccaggc cgacgaagggt gtcaccggcg cgcacccgaa 17280
ttcctttgtc ctgcaaatag ttccgtatc cgtcagcatc ggcgtatgg aacagtagcga 17340
aaggggccgc accatcgacc acctcgac ccaccgtatc cagtcggcc accatctccg 17400
cgccgcacgc cgtcaaccgc accgcacgtcg ctgcggcagc ggcacccgc cggggggcgc 17460
agcaaggcagc gatggccgtc agttgcaatg ttcccaacgg ccagtgcgtc cgctgcacgg 17520
tcaaccgagc cagcacgtct ggcgagccga ggcgttagcc cacccgcaat ccggccagcg 17580
accacgtttt cgtcaagcta cggagccacca gcacatcggtt cagcgagtca tcggccaacg 17640
attgcggctc gccgggaacc caatcagcga acgcctcgac gaccaccagg atgcgtcccg 17700

gccggcgtaa ctcgagcagc tgctcgccga ggtgcagcac cgaggtgggg ttggtcggat 17760
tacccacgac gacaaggctcg gcgtcgctcg gcacgtgcgc ggtgtccagc acgaacggcg 17820
gctttaggac aacatggtgc gccgtgattc cggcagcgct caaggctatg gccggctcgg 17880
tgaacgcggg cacgacgatt gctgcccgca cggacttag gttgtgcagc aatgcgaatc 17940
cctccgcgc cccgacgagc gggagactt cgtcacgggt tctgccatga cgttcagcga 18000
ccgcgtcttg cggccgggtgc acatcgctgg tgctcgata gcgggcccagc tccggcagca 18060
gcgcggcggag ctgcccggacc aaccattccg gggggccggtc atggcggacg ttgacggcga 18120
agtccagcac gccgggcgcg acatccgtat caccgtggta gcgcgcgcgc gcaagcgggc 18180
tagtgtctag actcgccaca gcgtaaaca gtagtggcc ggtgtgcggg ccaagaatcc 18240
agagcaccgcg cgacgcgttg tctacgcggc gacaaccgcg acatcacagg cagctaacag 18300
ggcgtcgccgcg gtgatgatcg tcaggccaag cagctgtgcc tgggcgatga gcacacggtc 18360
gaatggatgt cgatggtgat ccggaaagctc tgccgtgcgc agtgtgtgcg tggtaactg 18420
acagcggcga cgtgccgcag cggcgcattc gatcgggcac gtaagaagcc gatggctcg 18480
gcggcgggag cttgccgagg cggtagttga tcgcgcattc ccaggcactg gcggccgaca 18540
agagaatgt gttgcggacg tcctgaacaa tcgcccgtgt ttcgttgacg gcatccgcag 18600
ccaaacgtgg gtgtcgatga ggtagcgctt caccggtaa agcgttcgag cacgtcgct 18660
gacaacggag cgtccaaatc gtcgggcacg cggtacacgc catggtaat gcctaaccgc 18720
cgagtcctcat gaggatgcag cggcacaagc tttgcttacccg gtcgcgcgcg gcgggcaatc 18780
tcaacctctg cccgcccgtac acgagccgca gcagctcgga caggcgtgtc ttgcctcg 18840
gaacgcccac ccgcttcgca ggcgcggaga ctttgcgcgc gaccacctgc tcaccaaact 18900
tcgcgcattc cgcctgatac cacagcgcac acgggttagcg gtttgccttccaa ccgcttcg 18960
aacgacaatg ggatcgatcg cgacacgcacc gcgagcggga ccaattggcc gcctccctcca 19020
cgccgcgcgcg cacggcgccgc atcgctcgccg ggtgaatcgcc cgacgtggatcgat 19080
ctggacggca cgctgaccga ctcggcgccgc ggaatcgat ccagcttccg acacgcgcgc 19140
aaccacatcg gtgccccagt acccgaaggc gacctggcca ctcacatcgat cgccccgccc 19200
atgcgtgaga cgctgcgcgc catggggctc ggcgaatccg ccgaggaggc gatcgtagcc 19260
taccggggccg actacagcgc cccgcgttgg gcatgtgaaca gcttgcgtcg cggatcg 19320
ccgctgtgg ccgacacgtcg caccggcggt gtccggctgg ccgtcgccac ctccaaaggca 19380
gagccgaccg cacggcgaaat cctgcgcac ttcggaaattt agcagcactt cgaggtcatc 19440
gcggcgcgcgac gacccgttgg ctcggcgccgc agcaaggctcg acgtgtggc ccacgcgcgc 19500
gcgcagctgc ggccgctacc cgagcggttg gtgatggtgc ggcggccgcg ccacgacgtc 19560
gacggggccgg cccgcgcacgg catcgacacgc gtgggtggtgc gctggggcta cggcgccgc 19620

gactttatcg acaagacctc caccaccgtc gtgacgcattg ccggccacgt tgacgagctg 19680
agggaggcgc taggtgtctg atccgctgca cgtcacattc gtttgcacgg gcaacatctg 19740
ccggtcgcca atggccgaga agatgttcgc ccaacagctt cgccaccgtg gcctgggtga 19800
cgccgtgcga gtgaccagtg cgggcaccgg gaactggcat gttaggcagg tgccgcacga 19860
gcgggcggcc ggggtgttc gagccccacgg ctaccctacc gaccaccggg ccgcacaagt 19920
cgccaccgaa cacctggcg cagacctgtt ggtggcctt gaccgcaacc acgctcgct 19980
gttgcggcag ctccggcgtcg aagccgcccc ggtacggatg ctgcggtcat tcgacccacg 20040
ctcgggaacc catgcgctcg atgtcgagga tccctactat ggcgatcact ccgacttcga 20100
ggaggtcttc gccgtcatcg aatccgcctt gcccggcctg cacgactggg tcgacgaacg 20160
tctcgcgcgg aacggaccga gttgatgccc cgcctagcgt tccctgctgcg gcccggctgg 20220
ctggcgttgg ccctggcgtt ggtcgcgttc acctacctgt gctttacggt gtcgcgcgg 20280
tggcagctgg gcaagaatgc caaaaacgtca cgagagaacc agcagatcag gtattccctc 20340
gacaccccgc cggttccgct gaaaaccctt ctaccacagc aggattcgtc ggccggcggac 20400
gchgcaagtggc gcccgggtac ggcaaccggc cagttacccctc cggacgtgca ggtgctggcc 20460
cgactgcgcg tggggggaggc ggaccaggcg tttgaggtgt tggccccatt cgtggtcgac 20520
ggcggaccaa ccgtcctggc cgaccgtggc tacgtgcggc cccaggtggg ctcgcacgta 20580
ccaccgatcc cccgcctgca ggtcagacg gtgaccatca ccgcgcggct gcgtgactcc 20640
gaaccggacg tggggggcaa agacccatcc gtcagagacg gcttccagca ggtgtattcg 20700
atcaataccg gacaggtcgc cgcgctgacc ggagtccagc tggctgggtc ctatctgcag 20760
ttgatcgaag accaaccggc cgggctcgcc gtgctcgcc ttccgcacatct agatccggg 20820
ccgttcctgt cctatggcat ccaatggatc tcggttcggca ttctggcacc gatcggtttg 20880
ggcttatttcg cctacggcga gatccggcgc cgccgggggg aaaaagcggg gtcggccacca 20940
ccggacaaggc caatgacggc cgagcagaaa ctgcgtgacc gctacggccg ccggcggtaa 21000
accaacatca cggccaaatac cgcagcccccc gcctggacca cccgcgacag caccacggcg 21060
cgccgcagat cggccacccctt gggcgaccgg ccgtcgccca aggtggggccg gatctgcaac 21120
tcatgggttgtt accgggtggg cccacccagc cgacacgtcaa ggcggccagc aaacgcgc 21180
tcgacgcacac cggcggtggg gctggatgg cggcgccgt cgccgcgcga ggcccgtaacc 21240
gcaccgcggg gcgacccacc gaccaccggc gcgacatca ccaccagcac cgccgtcgcc 21300
cgtgcgcggcata ctagttggc ccagtcatcc aatcgtgctg cagcccaacc gaatcgagaa 21360
taacgcggcg agcggttagcc gatcatcgag tccagggtgt tgatggcacc atatccacg 21420
accgcaggca cgccgcgtcg aagccgcccac agcagcggca ccacctggc gtcggcgggtg 21480
ttttcgccca ccgactccag cggcgacacgc gtcaggcccc ggcggcccaag ctggggccggg 21540

tcacccccgc acagcgacgg cagcagccgt cgccgcgcct cgacatcgtc gcgttccaac 21600
aggccgata totggcggcc ggtgcgcgc agcgaagttc cgcccgacgc tgcccaggta 21660
gccgtcgccgg tggccgcccac gggccaggac ctgcccggta gccgtcgac tgccgcgcg 21720
agcaagccca cccgcgcgcac cagcaggccg acgtgtaccg caccggcgcac ccggccgtca 21780
cggttaggtga tctgctccag cttggcggcc gcccgaccga acagggccac cgatgacact 21840
cgtttgggt cgccgaacac gacgtcgacg aggccgcga tcagcacgac gacggccctg 21900
gtctgccagg tcgtatgcacaa cactccggca gcgtcgacaca cgtggtctac gctcagctat 21960
ttatgacctc atacggcagc tatccacgat gaagcggcca gctaccggg ttggcggacct 22020
gttgaacccg gcccgaatgt ttttgcggc agcgaatgtc atcatgcagc tggcagtgcc 22080
gggtgtcggg tatggcgtgc tggaaagccc ggtggacagc ggcaacgtct acaagcatcc 22140
gttcaagcgg gcccggacca cccggaccta cctggcggtg gcgaccatcg ggacggaaatc 22200
cgaccgagcg ctgatccggg gtggcggtga cgtcgccac cggcagggtc ggtcgacggc 22260
ctcgagccca gtgtcctata acgccttcga cccgaagttt cagctgtggg tggcggcgtg 22320
tctgtaccgc tacttcgtgg accagcacga gtttctgtac ggcccactcg aagatgccac 22380
cgccgacgcc gtctaccaag acgcacaaacg gtttagggacc acgctgcagg tgccggaggg 22440
gatgtggccg cccggacccgg tcgcgttcga cgagtactgg aagcgctcgc ttgatgggct 22500
gcagatcgac gcccgggtgc gcgagcatct tcgcgggggtg gcctcggtag cgtttctccc 22560
gtggccgttg cgcgcgggtgg cccggccgtt caacctgttt gcgacgacgg gattcttggc 22620
accggagttc cgcgcgatga tgcagctgga gtggtcacag gcccagcgc gtcgcgttcga 22680
gtggttactt tccgtgtac gtttagccga cccggctgatt ccgcacatcggtt cctggatctt 22740
cgtttaccag cttaacttggt gggacatgcg gtttcgcgcgc cgacacggcc gccaatcg 22800
ctgatagagc cccggccgagt gtgagcctga cagccccgaca cccggcggcgt gtgtcggtc 22860
gccaggttca cgctcggcga tctagagccg ccgaaaacct acttctgggt tgcctccgaa 22920
atcaacgtgc tgatctgctc gagcagctca cgcacatcggtt cgacatcgac atccacccgc 22980
gcatacagggt cggccttgggt cggccgcgc tggccgcacg tcattggccg caccggcgtt 23040
gctgtctgtc gcccggcgat gtcgtttga aacccaggta gctcacccac gaccacgaca 23100
ctgcacatatac cggcgcccccg ccgacaacgca agcacagcta gccgggtggc gcccgggg 23160
tcgaaccgcc gaccgctgggt gtgtaaaacc agagctctac cgctgagctca cccggccatg 23220
accggccgag gctacacgc ttgcggccaa gcacccaaaa ccttagggccg taagcgccgc 23280
cagagcgatcg gtccacagcc gtcgtatcgac aacttcaccc ggctgcttca tctcgccgaa 23340
ccgaatgatc cctgaccgat cgaccacaaa ggtgccccgg ttagcgatgc cggcctgctc 23400
gttgaagacg ccgttaggcct gactgaccgc gccgtgtggc cagaagtccg acaacagcgg 23460

aaacgtgaat ccgctctgcg tcgcccagat ctgtgagtg ggtggccggc ccaccgaaat 23520
cgctagcgcg gcgctgtcgt cggtctcaaa ctcggcagg tgatcacgca actggtccag 23580
ctcgccctgg cagatgccc tgaacccaa cgaaaagaac accaacagca cggtcttgc 23640
accccgtag cccgcgcagg tgacaagctg ctgattctgg tcgcgcaacg tgaagtcagg 23700
ggcgggtggct ccgacgttca gcatcagcgc ttgccagccc gcgatttcgg ctgtaccaat 23760
ctgctggcgc tccagttgcc cagattgacc gacgaggtcg gcatcagccc agctgtggc 23820
gccgcctcgg caatctcggc gggcaataca tgccgggct ggccggctt gggcgtcacc 23880
acccaaatca caccgtcctc ggcgagcggg ccgatcgcat ccatcagggt gtccaccaaa 23940
tcgcgcgtcgc catcacgcca ccacaacagg acgacatcga tgacctcgtc ggtgtcttca 24000
tcgagcaact ctcccccgca cgcttcttcg atggccgcgc ggatgtcg tcgttgtct 24060
tcgtcccagc cccattcctg gataagttgg tctcggttga tgcccaattt gcggcgttag 24120
ttcgaggcgt gatccgcgc gaccaccgtg gaacctcctt cagtctcgc gggccatgtg 24180
cacaccgtcg cgatggcat tatcgctcgca cagccagaac cggtccaccc gccccctca 24240
gaaggcggcc acgcacattt tcaatgcctt tgtcttggtg tcgttgagcc gatcaacccg 24300
ccgggtgaat tccgctgtcg acgcgtgcgc accgatggca tttgccaccg cgccggccgc 24360
gtcgacatat gcgtttagcg catccccag ttgcgcggac agcgcggcgc tcagactgcc 24420
tgagaccgtc gaggcactgt ttttgagcgc gtcgatggcc ggaccttcgg tcggcccggt 24480
gttgcggccc tgattgaacg cggccacgta ggcgttccacc ttgtcgatgg cgtccttgc 24540
ggtggccgccc agcgcgtcac acgaggtgcg aatgccttgc gtcgtcagcg attttggcg 24600
ctgcgactcc cggatgtcg acgtcgccgc cgaagccgac accgacgcgg acaccgacga 24660
gcggtaggcc ggtgcgacgt tggtgtcgcatggccgtta ccgtcggtga cagtggtaca 24720
tccgacgatc cccatcagca gcagcgcgtat gcagccgagc gccaggcgc ctcgccttgg 24780
gagctcccccc ccgtgcctgc gaggcacggc ggcgcattccg atgagcacgg catgtgaggt 24840
tacctggtcg cagcgcgacc ggcgtggccg tggtgtcgatcg cgcacccgca gaaccgagcg 24900
gagtgcggct atccgcccgc gacgcgggtg cggcacgata gggggacgac catctaaaca 24960
gcacgcgaagc ggaagccgc cacctacagg agtagtgcgt tgaccaccga ttgcggccgc 25020
cacgatctgg cccaaaactc aaacagcgcgca agcgaaccccg accgagttcg ggtgttccgc 25080
gaggggtgtgg cgtcgatattt gcccgcattt gatccccgagg agacctcgga gtggctggag 25140
tcctttgaca cgctgcgtca acgctgcggc ccgtcgccgg cccgctacact gatgttgcgg 25200
ctgctagagc gggccggcga gcagcgggtg gccatcccg cattgacgtc taccgactat 25260
gtcaacacca tcccgaccga gctggagccg tggttccccc ggcacgaaga cgtcgaacgt 25320
cgttatcgag cgtggatcag atggaatgcg gccatcatgg tgcaccgtgc gcaacgaccg 25380

ggtgtggcg tgggtggcca tatctgacc tacgcgtcg tccgcggcgct ctatgaggtc 25440
ggtttcaacc acttcttccg cgccaagtcg caccggcg gccgcgatca ggtgttcatc 25500
cagggccacg ctccccggg aatctacgacg cgccgccttc tcgaaggcg gttgaccgac 25560
gagcaactcg acggattccg ccaggaacac agccatgtcg gcggcggtt gccgtcctat 25620
ccgcacccgc ggctcatgcc cgacttctgg gaattccca ccgtgtcgat gggtttggc 25680
ccgctcaacg coatctacca ggcacggttc aaccactatc tgcatgaccg cggtatcaaa 25740
gacacccctcg atcaacacgt gtggtgcccc ttgggcgacg gcgagatgga cgaacccgag 25800
agccgtggc tggcccacgt cgccgcgctg gaaggcttgg acaacttgac cticgtgatc 25860
aactgcaatc tgcagcgact cgacggcccg gtgcgcggca acggcaagat catccaggag 25920
ctggagtcgt tcttccgcgg tgccggctgg aacgtcatca aggtggtgcg gggccgcgaa 25980
tggatgccc tgctgcacgc cgaccgcgac ggtgcgctgg tgaatttaat gaatacaaca 26040
cccgatggcg attaccagac ctataaggcc aacgacggcg gctacgtgcg tgaccacttc 26100
ttcggccgcg acccacgcac caaggcgctg gtggagaaca tgagcgacca ggatatctgg 26160
aacctcaaac gggcgccca cgattaccgc aaggtttacg ccgcctaccg cgccgcgc 26220
gaccacaagg gacagccgac ggtgatcccg gccaagacca tcaaaggcta cgcgctggc 26280
aagcatttcg aaggacgca tgccacccac cagatgaaaa aactgaccctt ggaagacctt 26340
aaggagtttc gtgacacgca gcggttcccg gtcagcgacg cccagttga agagaatccg 26400
tacctgcgc cctactacca ccccgccctc aacgccccgg agattcgta catgctcgac 26460
cgccgcgggg ccctcggggg ctgttcccg gagcgcagga ccaagtccaa agcgctgacc 26520
ctgccccgtc ggcacatcta cgccgcgcgtg aaaaagggtctt ggcacccatca ggaggtggcc 26580
accaccatgg cgacgggtcg cacgttcaaa gaagtgttcg ggcacaaagca gatcgccgc 26640
cgatagttcc cgatcattcc cgacgaggcc cgacccatcg ggatggactc ctggttcccg 26700
tcgctaaaga tctataaccg caatggccag ctgtataccg cggttgacgc cgacctgatc 26760
ctggcctaca aggagagcga agtcgggcag atcctgcacg agggcatcaa cgaagccggg 26820
tcgggtggct cggtcatcgcc ggccggcacc tcgtatgcga cgacacaacga accgatgatc 26880
cccatttaca tcttctactc gatgttccgc ttccagcgca ccggcgatag ctctggcc 26940
gcggccgacc agatggctcg agggttcgat ctcggggcca cccgcggcg caccaccctg 27000
accgggtgagg gcctgcaaca cgccgcacgt cactcgatcg tgctggccgc caccaccccg 27060
gcgggtgggt cctacgaccc ggccctcgcc tacgaaatcg cctacatcgat ggaaagcgaa 27120
ctggccagga tgtgcggggaa acacatcttct tctacatcac cgtctacaac 27180
gagccgtacg tgcagccgac ggagccggag aacttcgatc ccgaggcgat gctgcggggat 27240
atctaccgt atcacgcggc caccgagcaa cgccaccaaca aggcgcagat cctggccctcc 27300

ggggtagcga tgcccgcggc gctgcgggca gcacagatgc tggccgccga gtggatgtc 27360
gccgccgacg tgtggtcggt gaccagttgg ggcgagctaa accgcgcacgg ggtggccatc 27420
gagaccgaga agctccgcca ccccgatcgg cccgcggcg tgccctacgt gacgagagcg 27480
ctggagaatg ctcggggccc ggtgatcgcg gtgtcgact ggatgcgcgc ggtccccgag 27540
cagatccgac cgtgggtgcc gggcacatac ctcacgttgg gcacccgacgg gttcggctt 27600
tccgacactc ggcccgcgc tcgcccgtac ttcaacaccg acgcccgaatc ccaggtggtc 27660
gcgggtttgg aggcggtggc gggcgacggc gagatcgacc catcggtgcc ggtcgccggc 27720
gcccgcagt accggatcga cgacgtggcg gctgcgcccggc agcagaccac ggatcccggt 27780
cccgccccct aacgcggcg agccgaccgc ctttggccga atcttccaga aatctggcgt 27840
agcttttagg agtgaacgac aatcagttgg ctccagttgc cccgcggagg tcgcccgtcg 27900
aactgctgga cactgtgccc gattcgctgc tgccgcgggtt gaagcagttac tcggccggc 27960
tggccaccga ggcagttcg gccatgcaag aacggttgcc gttcttcgccc gacctagaag 28020
cgtcccaagcg cgccagcgtg ggcgtgggtt tgccagacggc cgtggtaaac ttctcgat 28080
ggatgcacga cccgcacagt gacgtcggtt ataccgcga ggcattcgag ctggtcccc 28140
aggatctgac gcgacggatc gcgcgtcgcc agaccgtgga catggtgcgg gtcaccatgg 28200
agttttcgaa agaagtctgtt cccctgtcg cccgttccga agagcagttt accggccctca 28260
cggtggcat tttgaaatac agccgcgacc tggcattcac cgccgcacg gctacgcgg 28320
atgcggccga ggcacggaggc acctggaca gccggatggaa ggccagcgtg gtggacgcgg 28380
tggtaacgcgg cgacaccggc cccgagctgc tggccggggc ggccgcgtt aattgggaca 28440
ccaccgcgccc ggccggcgta ctggggaa ctccggcgcc cggccaaat ggctccaaaca 28500
gcgacggcga cagcgagcgg gccagccagg atgtccgcga caccgcggct cgccacggcc 28560
gcgcgtcggtt gaccgcgtt cacggcacct ggctgggtt gatcgctcc ggccagctgt 28620
cgccaaaccga gaagttccctc aaagacctgc tggcagcatt cgccgacgccc ccgggtggta 28680
tcggccccac ggccggccatg ctgaccgcgg cgccacccgcg cgcgtcgag gogatctccg 28740
ggatgaacgc cgtcgcggc tggcgcggag cgccgcggcc cgtgcgtggct agggacttt 28800
tgcccgaaacg cgccctgtatg ggccgcgtt cggcgatcgt ggccctgtcat accgacgtga 28860
tgcggccccc agccgatgcc ggaccgcgc tcatcgagac gctagacgcata tatctggatt 28920
gtggcggcgc gattgaagct tggccagaa agttgttcgt tcatccaaac acagtgcgg 28980
accggctcaa gcggatcacc gacttcaccg ggccgcgttcc caccgcgcac cgcgtgcct 29040
atgtccctcg ggtggcggcc accgtgggtc aactcaacta tccgacgcgc cactgaagca 29100
tcgacagcaa tgccgtgtca tagattccct cgccggcgtt aggggttcca gcagggccccc 29160
cgaaagata ccaggggcgc cgtcgacgg aaagtgtatcc agacaacagg tcgccccacg 29220

atctcaaaaa catagcttac aggcccgtt ttttttttat atacaaaaac ctaagacgag 29280
gttcataatc tgttacaccg cgcaaaaccg tttcacagt gttcttttag acacgtgatt 29340
gcgttgctcg cacccggaca gggttcgcaa accgaggaa tttgtcgcc gtggcttcag 29400
ctgcccggcg cagcggacca gatcgcggcg tggcgaaag ccgctgatct agatcttgcc 29460
cggctggca ccaccgcctc gaccgaggag atcaccgaca ccgcggtcgc ccagccattg 29520
atcgctcgccg cgactctgct gcccaccag gaactggcgc gccgatgcgt gctgcccggc 29580
aaggacgtca tcgtggccgg ccactccgtc ggcgaaatcg cggcctacgc aatcgccgg 29640
gtgatagccg ccgacgacgc cgtcgcgtg gccgcaccc gcccgcggc gatggccaag 29700
gcctgcgcca ccgagccgac cggcatgtct ggggtgtcg gcccgcacga gaccgagg 29760
ctgagtcgccc tcgagcagct cgacttggtc cccgcaaaacc gcaacgcgcg cggccagatc 29820
gtcgctgccc gccggctgac cgcgttggag aagctcgccg aagaccggcc ggccaaggcg 29880
cgggtgcgtg cactgggtgt cccggagcg ttccacaccg agttcatggc gcccgcactt 29940
gacggctttg cggcggccgc ggccaacatc gcaaccgcgc accccaccgc cactgtgt 30000
tccaaccgcg acgggaagcc ggtgacatcc gccggccgcgg cgatggacac cctggcttcc 30060
cagctcaccc aaccgggtcg atgggacctg tgcaccgcga cgctgcgcga acacacagtc 30120
acggcgatcg tggagttccc ccccgccggc acgcttagcg gtatcgccaa acgcaactt 30180
cgggggggttc cggcacgcgc cgtcaagtca cccgcagacc tggacgagct ggcaaaccta 30240
taaccgcgga ctggccaga acaaccacat acccgtcaat tcgatttgcg cacaacat 30300
tacgaaggga agcatgtgt gcctgtcaat caggaagaaa tcattggccg tattggccgag 30360
atcatcgaaag aggtAACGGG tatcgagccg tccgagatca cccggagaa gtcgttcgtc 30420
gacgacctgg acatcgactc gctgtcgatg gtcgagatcg ccgtgcagac cgaggacaag 30480
tacggcgatca agatccccga cgaggacctc gccggctctgc gtaccgtcg tgacgttgc 30540
gcctacatcc agaagctcgaa ggaagaaaac ccggaggccgg ctcaggcggt gcccgcgaag 30600
attgagtcgg agaaccggaa tgccgttgcc aacgttcagg cgaggcttga ggccgagtcc 30660
aagtgagtca gcattccacc gctaattggcg gtttccccag cggtgtggg accggccgtca 30720
cagcgacgac gtcgatctcg ccggacatcg agagcacgtg gaagggtctg ttggccggcg 30780
agagcggcat ccacgcactc gaagacgagt tcgtcacca gtcggatcta gcccgtcaaga 30840
tcggcggtca cctcaaggat ccggcgatc gacccatggg ccgactcgac atgcgacgca 30900
tgtcgatcg ccagcggatg ggcaagttgc tggcgccgaca gctatgggag tccggccggc 30960
gcccggaggt cgatccagac cggatccgcg ttgttgcgg caccggatca ggtggagccg 31020
agaggattgt cgagagatc gacctgtatga atgcggggggg ccccccggaaag gtgtccccgc 31080
tggccgttca gatgatcatg cccaaacggtg ccgcggccggat gtcggatcg cagcttgggg 31140

cccgccgg ggtgatgacc ccgggtcggt cctgttcgtc gggctcgaa gcgatcgccc 31200
acgcgtggcg tcagatcggt atgggcgacg ccgacgtcg cgtctcgcc ggtgtcgaaag 31260
gaccatcgaa ggcgctgccc atcgccgggt tctccatgtat gcggggccatg tcgacccgca 31320
acgacgagcc tgagcgggccc tcccgccgt tcgacaagga ccgcgacggc tttgtgttcg 31380
gcgaggccgg tgcgctgatg ctcatcgaga cggaggagca cgccaaagcc cgtggcgcca 31440
agccgttggc ccgattgctg ggtgccggta tcacctcgga cgccttcat atggtggcg 31500
ccgccccgaa tgggtgtcgatg gccggtaggg cgatgactcg ctcgctggag ctggccgggt 31560
tgtcgccggc ggacatcgac cacgtcaacg cgcacggcac ggcgacgcct atcggcgacg 31620
ccgcggaggc caacgcccattc cgcgtcgccg gttgtatca ggccgcggtg taacgcggca 31680
agtctgcgtt gggccactcg atcggcgccg tgggtgcgt cgagtcggtg ctacgggtgc 31740
tgacgctgcg cgcggcgatc atcccgccgaa ccctgaacta cgagacaccc gatcccgaga 31800
tcgaccttga cgtcgccgccc ggcgaaccgc gctatggcga ttaccgtac gcagtcaaca 31860
actcggttgcg gttcgccggc cacaatgtgg cgcttgcctt cggcggttac tgaagcacga 31920
catcgccggatcg cgcggggccc gaggtgggggg tccccccgt tgcggggcg agtcggaccg 31980
atatgaaagg aacgttcgca agaccaatga cggagctgg taccggaaa gcctttccct 32040
acgttagtcgt caccggcattc gccatgacga ccgcgtcg caccgacgcg gagactacgt 32100
ggaagttgtt gctggaccgc caaagcgaa tccgtacgt cgatgaccca ttctcgagg 32160
agttcgacccgtt gccagttcgatc atcggcgac atctgcttgcg ggaattcgac caccagctga 32220
cgccggatcgaa actgcggccgg atgggatacc tgcagcgat gtccaccgtg ctgagccggc 32280
gcctgtggaa aatgcggc tcacccgagg tggacaccaa tcgattgtatgcgttccatcg 32340
gcacccggcct ggggtcgccg gaggaactgg ttttcgttta cgacgtatgcgtcccg 32400
gaatgaaggc ggtctcgccg ctgaccgtgc agaagtacat gccaacggg gccgcccgg 32460
cggtcggtt ggaacggcac gccaaggccg ggggtatgcgttcccgatcgatcg 32520
ccggccgcga gcccattcgcc cgtcggtggc agcagattgt gctggagag gccgatgccc 32580
ccatctcgccg cggcggtggag accaggatcg aagcggtgcc catcgccggg ttctcgatcg 32640
tgcgcacatcgatgttccacc aacaacgacg accccggccgg tgcacgtccgc ccattcgaca 32700
gggaccgcga cggcttgcgtt ttcggcgagg gggcgccct tctgttgcgtt gggaccgg 32760
agcacgcac ggcacgtggc gccaacatcc tggccggat catggcgcc agcatcacct 32820
ccgatggctt ccacatgggt gccccggacc ccaacggggaa acgcgcgggg catgcgatcg 32880
cgccggcgat tcagctggcg ggcctcgccc cggcgacat cgaccacgtc aatgcgcacg 32940
ccacccggcac ccaggtcgcc gacctggccg aaggcaggccatcaacaac gccttggcg 33000
gcaacccgacc ggcgggtgtac gcccccaagt ctgcctcggtt ccactcggtg ggcgggtcg 33060

gcgcggtcga atcgatcttg acggtgctcg cggtgcgcga tcaggtgatc ccggccgacac 33120
tgaatctggt aaacctcgat cccgagatcg atttggacgt ggtggcgggt gaaccgcgac 33180
cgggcaatta cccgtatgcg atcaataact cgttcggatt cggcggccac aacgtggcaa 33240
tcgccttcgg acggtaactaa accccagcgt tacgcgacag gagacctgctg 33300
tggcccccga ggcgggttggc gagtcgctcg acccccgcga tccgctgttg cggctgagca 33360
acttcttcga cgacggcagc gtggattgc tgacgcgacg tgaccgcgtcc ggagtgcgtgg 33420
ccgcggcggg caccgtcaac ggtgtgcgca ccatcgcgtt ctgcaccgcac ggcaccgtga 33480
tgggcggcgc catgggcgtc gaggggtgca cgacacatcgta caacgcctac gacactgcca 33540
tcgaagacca gagtcccatac gtgggcatac ggcatcggtt tggtgcccggtt ctggctgaag 33600
gtgtgcgggc gctgcacgcg gttaggcagg tggcgtggc catgatccgc gcgtccggct 33660
acatccgcga gatctcggtt gtcgtcggtt tcgcccggcggc tacggaccgg 33720
cggtgaccga cgtcgtcggtc atggcggcggaa aagccgggtt gttcgtcacc gggcccgacg 33780
tggtgccgcg cgtcaccggc gaggacgtcg acatggcctc gtcgggtggg ccggagaccc 33840
accacaagaa gtccgggggtt tgccacatcg tcgcccggcga cgaactcgat gcctacgacc 33900
gtggcggccg gttggtcggtt ttgttctgccc agcagggggca tttcgatcgc agcaaggccg 33960
aggccgggtga caccgacatc cacgcgtgc tgccggaaatc ctcgcgacgt gcctacgacg 34020
tgcgtccgtatc cgtgacggcg atcctcgatc cgacacacacc gttcgtacgag ttccaggcc 34080
attggcgcgc gtcgtatgggtt gtcgggttgg gtcggctgtc gggtcgcacg gtgggtgtac 34140
tggccaaacaa cccgctacgc ctggggcgct gcctgaactc cgaaagcgca gagaaggcag 34200
cgcgtttcgatc gcccgggtgc gaccaggagt ggggtggcgtt ggtgcggccgtt ggcggccaaatc 34260
cggttatctt gcccgggtgc gaccaggagt ggggtggcgtt ggtgcggccgtt ggcggccaaatc 34320
tgctgcacgc gttcggcgcag tgcaccgttc cgcgggtcac gtcgggttacc cgaaagaccc 34380
acggcggggc atacattcgatc atgaactccc ggtcggttggaa cgcgaccaag gtgttcgcct 34440
ggccggacgc cgaggatcgatc gtgtatggcgat ctaaggcggc cgtcggcatac ctgcacaaga 34500
agaagttggc cgccgctccg gagcacgaac gcaagcgct gcacgaccag ttggccggcc 34560
agcatgagcg catcgccggc ggggtcgaca gtgcgttggaa catcggtgtt gtcgacgaga 34620
agatcgaccc ggcgcataact cgcagcaagc tcaccggcgt gtcggcgcac gtcggccac 34680
ggcgcggccg ccacaagaac atcccgctgtt agttctgacc gtcggcgcac gtcggccac 34740
acgcgcgagg tccgcggccgt gcgattctgc gtcgtgcgc cagttatccc cagcggtggc 34800
tggtaaacgc gaggcgctcc tcgcgtgc cggcgttggcc taccgacgcg ctaacaattc 34860
tcgagaaggc cggcgggttc gccaccaccg cgcattgtt ctcggcgtatc acccgccaaac 34920
agctcgacgt ccaagtgaaa aacggcggcc tcgttcgcgtt tggtaacggc gtcgtacgcgg 34980

cacaagagcc ggacctgttg ggccgcttgg cggctctcg a tgtgttcatg ggggggcacg 35040
ccgtcgcggt tctgggcacc gcccggcggt tgtatggatt cgacacggaa aacaccgtcg 35100
ctatccatat gctcgatccc ggagtaagga tgcggccac ggtcggtctg atggtccacc 35160
aacgcgtcgg tgccggctc caacgggtgt caggtcgctc cgcgaccgcg cccgcatgga 35220
ctgcccgttga ggtcgacgca cagttgcgcc gcccggggc gctggccacc ctcgacgccc 35280
caactacggtc aatgcgctgc gctcgcaagt aaattgaaaa cgccgttgct gagcagcggag 35340
gcccggagg catcgctcg ggcgcgcaac tcttaccctt cgccgacgga cgcgccgaaat 35400
cgccccatgga gagcgaggct cggctcgta tgcacgatcca cgggctgccc ttgcccgaac 35460
ttcaataccca gatacacggc cacggtggtg aaatgtggcg agtcgacttc gcctggcccg 35520
acatgcgtct cgccggcga tacgaaagca tcgagtgga cgcgggaccg gcggagatgc 35580
tgcgcgacaa gacacgctgg gccaagctcc aagagctcggt gtggacgatt gtcccgattg 35640
tcgtcgacga tgtcagacgc gaaccggcc gcctggcgcc cgcacatcgcc cgccaccccg 35700
accgcgcgca tatggccggc tgaccgctgg tgagcagacg cagagtcgca ctgcccgggg 35760
cgcaagtgcga ctctgcgtct gctcgctc aacggctgag gaactcctta gcccacggcga 35820
ctacgcgtc gcatccccgt ggcaccagac cgatccgggt cggccggcgtc aggatatacg 35880
ccacatccag cgccccctca tgggtcaccc cgtattcgaa ctccggccgg gtcacgtcga 35940
tgccgtcgcc gaccggctcg gtggggccct cacatgtggc ggcggcagcg acgttggccg 36000
cctcgccccc gtaccgcgccc accagcgact cgggcaatcc ggcggccgat ccggggccgg 36060
gcccagggtt cgccgggtcg ccgatcagcg gcaggttgcg agtgcggcac ttgcggctc 36120
gcaggtgtcg cagcgtatcg ggcgcattca gcacatcc tcgcattgttag cggtattccg 36180
tcagcttgc gccgaccaca ctgatcacgc ccgacggcga ttcaaaaaca gcgtggtcac 36240
gcgaaacgtc ggccgggtcg ccctggacac cagcaccggc ggtgtcgatt agcggccgca 36300
atccccgcata ggcaccgatg acatccttgg tgccgaccgc cgtccccat gcggtgttca 36360
ccgtatccag caggaacgtg atctcttccg aagacgggtt tgccacatcg ggaatcgggc 36420
cgggtgcggtc ttgcgtcggtc agccccgagat agatccggcc cagctgctcg ggcattggcga 36480
acacgaagcg gttcagctca cgggggatcg gaatggtcag cgcggcagtc ggattggcaa 36540
acgacttcgc gtcaagacc agatgtgtgc cgcggctggg gcgtagccctc agggacgggt 36600
cgatctoacc cgcggccacacg cccgcccgt tgatgacggc acgcggccgac agcgcgaacg 36660
actgcccgggt gcggccgggtcg gtcaactcca ccgaagtgc ggtgacattc gacgcggccca 36720
cgtaagttag gatgcggccg ccgtgttggg cgcgggtcg cgcgacggcc atgaccagcc 36780
gggcgtcggtc gatcaattgc ccgtcgta cgcggccgac accgtcgagg ccgtccccggcc 36840
gaacgggtggg agcaatctcc accaccgggt acgcggggat tcggcgcgat cggggcaacg 36900

tcggccggcgg cgtacccgct agcacccgca aagcgtcgcc ggccaggaaa cccgcacgca 36960
 ccaacgccccg cttggtgtga cccatcgacg gcaacaacgg gaccagttgc ggcatggcat 37020
 gcacgagatg aggagcggtt cgtgtcatca ggattcccg cgatgcccac gttgccgctg gccagatagc gcagaccgcc gtgcaccaac ttcgagctcc 37080
 agcggctggt gccgaacgcc agatcatgct tttccaccaa ggccaccgtc agaccgcggg 37140
 tggcagcattc taaggcaatg ccaacaccgg taatgccgcc gcctatcacg atgacgtcga 37260
 gtgcgccacc gtcggccagt gcggtcaggt cgccggagcg acgcgcgcgc ttgagtgcag 37320
 ccgagtgggg catcagcaca aatatccgtt cagtgcgtgg gtaagttcgg tggccagcgc 37380
 ggccgaatcg aggatcgaat cgacgatgtc cgccggactgg atggtcgact gggcgatcag 37440
 caacaccatg gtcggccagt gacgagcgac gccggagcgc acactgcccgc accgctgcgc 37500
 cactgtcagc cggccggcca acccctcgat caggacctgc tggctgggtgc cgaggcgctc 37560
 ggtatgtac accctggcca gctccgagtg catgaccgac atgatcagat cgtcaccccg 37620
 caaccggtcg gccaccgcga caatctgctt taccacacgt tcccggtcg ccccgatcg 37680
 gggcacctcc cgccgacgt cgccgatatg gctggtcagc atggacgcga tgatcgaccg 37740
 ggtgtccggc cagcgacggt atacgggtcg gccgctcactc cccgcgcgc gggcgatctc 37800
 ggcaagtgtc acccggtcca cgccgtaatc gacgacgcag ctcggcgctg cccgcaggat 37860
 acgaccaccg gtatccgcgc ggtcattact cattgacagc atgtgtataa ctgttaacgog 37920
 tgactcaccg cgaggaactc cttccaccga taaaatggg cgcgtgggg gatccgcgcg 37980
 cggccaagcc actttctgat ggcgtccggc cggtgctgaa gcagggtgtg ggcctagcgg 38040
 actcggagca gccccgactc gaccccgccg aggtgcagct ggcggcgcc gccctgtcgg 38100
 gggcagacca 38110

<210> 25

<211> 2540

<212> DNA

<213> Homo sapiens

<400> 25
 gaaaagggtgg acaagtccata ttttcaagag aagatgactt ttaacagttt tgaaggatct 60
 aaaacttggc tacctgcaga catcaataag gaagaagaat ttgtagaaga gtttaataga 120
 ttaaaaactt ttgctaattt tccaaatgggt agtccctgttt cagcatcaac actggcacga 180
 gcagggtttc ttataactgg tgaaggagat accgtgcggg gcttttagtt tcatgcagat 240
 gtagatagat ggcaatatgg agactcagca gttggaagac acaggaaagt atccccaaat 300

tgcatgtttt tcaacggctt ttatcttgc aatagtgcac cgcatgtctac aaattctgg 360
atccagaatg gtcagtacaa agttgaaaac tatctggaa gcagagatca ttttgcctt 420
gacaggccat ctgagacaca tgcagactat cttttgagaa ctggggcaggt tggatata 480
tcagacacca tatacccgag gaaccctgcc atgtattgtg aagaagctg attaaagtcc 540
tttccagaact gcccagacta tgctcaccta accccaagag agtttagcaag tgctggactc 600
tactacacag gtattggtga ccaagtgcag tgctttgtt gtgggtggaaa actgaaaaat 660
tgggaacctt gtgatcgtgc ctggtcagaa cacaggcgac actttcctaa ttgtttttt 720
gttttggcc ggaatcttaa tattcgaagt gaatctgatg ctgtgagttc tgataggaat 780
ttcccaaatt caacaaatct tccaagaaat ccatccatgg cagattatga agcacggatc 840
tttacttttggacatggat atactcagtt aacaaggagc agcttgcaag agctggattt 900
tatgcttttag gtgaagggtga taaagtaaag tgctttcact gtggaggagg gctaaactgat 960
tggaagccca gtgaagaccc ttgggaacaa catgctaaat ggtatccagg gtgcaaataat 1020
ctgttagaac agaagggaca agaatatata aacaatattc attaaactca ttcaacttgag 1080
gagtgtctgg taagaactac tgagaaaaca ccatcactaa ctagaagaat tgatgatacc 1140
atcttccaaa atcctatggt acaagaagct atacgaatgg ggttcagttt caaggacatt 1200
aagaaaataa tggaggaaaaa aattcagata tctggagca actataatc acttgaggtt 1260
ctgggtgcag atcttagtga tgctcagaaa gacagtatgc aagatgagtc aagtcagact 1320
tcattacaga aagagattag tactgaagag cagctaaggc gcctgcaaga ggagaagctt 1380
tgcaaaatct gtatggatag aaatattgct atcgttttt ttccttgg acatctagtc 1440
acttgtaaac aatgtgtga agcagggtgac aagtgtccca tgtgctacac agtcattact 1500
ttcaagcaaa aaatttttat gtcttaatct aactctatag taggcatgtt atgttggct 1560
tattaccctg attgaatgtg tgatgtgaac tgactttaag taatcaggat tgaattccat 1620
tagcatttgc taccaagtag gaaaaaaaaat gtacatggca gtgttttagt tggcaatata 1680
atctttgaat ttcttgattt ttcagggtat tagctgtatt atccatttt tttactgtta 1740
tttaattgaa accatagact aagaataaga agcatcatac tataactgaa cacaatgtgt 1800
attcatagta tactgatttta atttctaagt gtaagtgaat taatcatctg gatTTTTT 1860
tcttttcaga taggcttaac aaatggagct ttctgtatataa atgtggag attagagttt 1920
atctccccaa tcacataatt tgtttggat gaaaaaggaa taaattttc catgctgg 1980
gaaagataga gattttttt agaggtgggt tggatgttt taggattctg tccatTTCT 2040
tgtaaaggaa taaacacgga cgtgtgcgaa atatgtttgt aaagtgtttt ggcattgtt 2100
aaagcgtatt taatgataga atactatcga gccaacatgt actgacatgg aaagatgtca 2160
gagatgtttt aagtgtaaaa tgcaagtggc gggacactat gtatgtctg agccagatca 2220

aagtatgtat gttgttaata tgcatagaac gagagatgg gaaagatata caccaaactg 2280
 ttaaatgtgg tttctcttcg gggagggggg gattggggga ggggccccag aggggtttta 2340
 gaggggcctt ttcaactttcg actttttca tttgttctg ttccggatttt ttataagtat 2400
 gtagaccccg aagggtttta tgggaactaa catcagtaac ctaaccccg tgactatcct 2460
 tgcctcttcc tagggagctg tgggtttcc caccaccac cttccctct gaacaaatgc 2520
 ctgagtgctg gggcactttg 2540

<210> 26

<211> 103

<212> RNA

<213> Homo sapiens

<400> 26
 agcuccuaua acaaaagucu guugcuugug uuucacauuu uggauuuccu aauauaaugu 60
 ucucuuuuua gaaaaggugg acaaguccua uuuucaagag aag 103

<210> 27

<211> 28

<212> RNA

<213> Homo sapiens

<400> 27
 ggauuuccua auauaauguu cucuuuuu 28

<210> 28

<211> 1619

<212> DNA

<213> Homo sapiens

<400> 28
 ccgcaggatt tgaatcgccg gaccgttgg cagaggtggc ggcggccgc tgggtgc 60
 gacgttgc 120
 gacgttgc cctgcctggc agcccttct caaggaccac cgcatctca cattcaagaa 120
 ctggcccttc ttggagggtc gcgcctgcac cccggagcgg atggccgagg ctggcttcat 180
 ccactgcccc actgagaacg agccagactt ggcccagtgt ttcttctgct tcaaggagct 240
 ggaaggctgg gagccagatg acgaccccat agaggaacat aaaaagcatt cgtccggttg 300

cgcttcctt tctgtcaaga agcagttga agaattaacc cttggtaat tttgaaact 360
 ggacagagaa agagccaaga acaaaattgc aaaggaaacc aacaataaga agaaagaatt 420
 tgaggaaact gogaagaaag tgccgtgc catcgagcag ctggctgcca tggattgagg 480
 cctctggccg gagctgcctg gtcccagagt ggctgcacca cttccagggt ttattccctg 540
 gtgccaccag cttccctgtg ggcccttag caatgtctta ggaaaggaga tcaacatTTT 600
 caaatttagat gtttcaactg tgctcctgtt ttgtcttga agtggcacca gaggtgcttc 660
 tgcctgtgca gcgggtgctg ctggtaacag tggctgcttc tctctctc tctttttt 720
 ggggctcat ttttgctgtt ttgattcccg ggcttaccag gtgagaagtg agggaggaag 780
 aaggcagtgt ccctttgct agagctgaca gctttgtcg cgtggcaga gccttccaca 840
 gtgaatgtgt ctggacctca tggatgtgag gctgtcacag tcctgagtgt ggacttggca 900
 ggtgcctgtt gaatctgagc tgcaggttcc ttatctgtca cacctgtgcc tcctcagagg 960
 acagttttt tggatgtgt tttttttttt tttttttttt ggttagatgca tgacttggat 1020
 gtgatgagag aatggagaca gaggccctgg ctccctact gtttaacaac atggcttct 1080
 tattttgtt gaattgttaa ttcacagaat agcacaaact acaattaaaa ctaagcacaa 1140
 agccattcta agtcattggg gaaacggggta gaaacttcagg tggatgagga gacagaatag 1200
 agtataatggg agcgtctggc agataactcct tttgccactg ctgtgtgatt agacaggccc 1260
 agtgcgcgc gggcacatg ctggccgctc ctccctcaga aaaaggcagt ggctaaatc 1320
 ctttttaat gacttggctc gatgtgtgg gggactggct gggctgtgc agggctgtg 1380
 tctgtcagcc caacccacac atctgtcacg ttctccacac gggggagaga cgcaatccgc 1440
 ccaggcccccc gtttctttt gaggcagcag ctccctcagg gctgaagtct ggctaaat 1500
 gatggatTTT attcccttc ctccctgtca tagagctgca gggggattt ttacagcttc 1560
 gctggaaacc tctggagggtc atctcggctg ttccctgagaa ataaaaagcc tgcatttc 1619

<210> 29

<211> 27

<212> RNA

<213> Homo sapiens

<400> 29

ggcguacacac cuucggguga agucgcc

27

<210> 30

<211> 27

<212> RNA

<213> Homo sapiens

<400> 30

ggcgucacac cuucggguga agucgcc

27

<210> 31

<211> 12

<212> PRT

<213> Homo sapiens

<400> 31

Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro
1 5 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/11757

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68; C07H 21/02; G01N 27/26
US CL : 435/6; 536/23.1; 204/451

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6; 536/23.1; 204/451

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
STN, EAST

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6329146 B1 (Crooke et al) 11 December 2001 (11.12.2001), column 40, example 11.	1
Y	US 5,807,682 A (Grossman et al) 15 September 1998 (15.09.1998), column 19, lines 2-18. <i>61</i>	1
Y	US 6,355,428 (Schroth et al) 12 March 2002 (12.03.2002), column 8, lines 64-67.	1
Y	US 6,320,040 B1 (Cook et al) 20 November 2001 (20.11.2001), column 11, lines 14-22	1
Y	US 6,391,542 B1 (Anderson et al) 12 May 2002 (12.05.2002), column 36, example 18.	1

<input type="checkbox"/>	Further documents are listed in the continuation of Box C.	<input type="checkbox"/>	See patent family annex.
*	Special categories of cited documents:		
"A"	document defining the general state of the art which is not considered to be of particular relevance	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"B"	earlier application or patent published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means	"Z"	document member of the same patent family
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 22 June 2002 (22.06.2002)	Date of mailing of the international search report 18 JUNE 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Jyothsna Venkat Signature: <i>Jyothsna Venkat Ph.D.</i> Telephone No. (703) 305-3235 EXAMINER TECHNOLOGY CENTER 1600